Whole Genome Analysis and Biomarker Identification of *Leptospira spp.*

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig zur

Erlangung des Grades einer

Doktorin der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Siti Roszilawati Binti Ramli
aus Negeri Sembilan

1. Referent: apl. Professor Dr. Michael Hust
2. Referent: Professor Dr. Michael Steinert
eingereicht am: 16.11.2018
mündliche Prüfung (Disputation) am: 10.01.2019

Druckjahr 2019

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor ***** der Arbeit,
in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Ramli, SR., Moreira, GMSM., Zantow, J., Goris, MGA., Nguyen, VK., Novoselova,N., Pessler, F. & Hust, M. (2019). Discovery of *Leptospira spp.* seroreactive peptides using ORFeome phage display. PLoS Negl.Trop.13(1) : e0007131.

Tagungsbeiträge

Ramli, SR., Bunk, B., Strommenger, B., Goris, MGA., Geffers, R., Pessler, F. Whole genome sequences analysis of *Leptospira interrogans* from clinical source (Vortrag) 10th International Leptospira Society Conference, Palmerston North, New Zealand (2017).

Ramli, SR., Moreira, GMSM., Zantow, J., Pessler, F., Hust., M. Identification of Novel Immunogenic Proteins of *Leptospira spp.* Using ORFeome Phage Display (Vortrag) 2nd International Conference of Molecular Diagnostic and Biomarker Discovery, Penang, Malaysia (2017).

Ramli, SR., Strommenger, B., Bhuju, S., Jarek, M.,Geffers, R., Pessler, F. Whole genome sequence based typing *Leptospira interrogans* from Malaysia (Poster). Applied Bioinformatics & Public Health Microbiology, Cambridge, United Kingdom (2017).

Ramli, SR., Bhuju, S., Jarek, M., Strommenger, B., Geffers, R., Pessler, F. Whole genome analysis of *Leptospira interrogans* Variant from Clinical Source: A Pilot Study (Poster) 26th European Congress of Clinical Microbiology and Infectious Diseases, Amsterdam, the Netherlands (2016).

TABLE OF CONTENTS

List of Abbreviations

AFAS	albumin fatty acid supplement
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AP-PCR	arbitrarily primed PCR
APTT	activated partial thromboplastin time
AST	aspartate transaminase
ATCC	American Type Culture Collection
bp	basepair(s)
BSA	bovine serum albumin
CAAT	cross agglutination absorption test
CFT	complement fixation test
CFU	colony-forming unit
cgMLST	core genome multilocus sequence typing
CK	creatinine kinase
CRISPR	clustered regularly interspaced short palindromic repeats
CRP	C-reactive protein
Ct	cycle threshold
CFU	colony forming unit
DDH	DNA-DNA hybridization
DFM	dark field microscope
DNA	deoxyribonucleic acid
ECM	extracellular matrix
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
EMJH	Ellinghausen-McCullough-Johnson-Harris
g	gram
GGDC	genome to genome distance calculator
h	hour
H&E	hematoxylin & eosin
HGT	horizontal gene transfer
HR	homologous recombinant
HTAR	Hospital Tuanku Ampuan Rahimah
ICU	intensive care unit
IFA	immunofluorescence assays
IFN	interferons
IHA	indirect hemagglutination assay
IgG	immunoglobulin G
IgM	immunoglobulin M
IL	interleukin
ILI	influenza-like illness

IMAC	immobilized metal ion affinity chromatography
INR	international normalized ratio
IPTG	Isopropyl-β-D-thiogalactopyranosid
IS	insertion sequences
IV	intravenous
kb	kilobase(s)
kDa	kilodalton(s)
kV	kilovolt(s)
L	liter
LCB	large collinear block
LFA	lateral flow assay
M	molar
mA	milliampere(s)
mAB	monoclonal antibody
MALDI-TOF	matrix-assisted laser desorption/ionization–time-of-flight
MAT	microscopic agglutination test
MCAT	microcapsule agglutination test
mg	milligram
MGE	mobile genetic element
min	minute
mL	milliliter(s)
MLST	multilocus sequence typing
mM	millimolar
mmHg	millimeter mercury
mRNA	messenger RNA
MOI	multiplicity of infection
MS	mass spectrometry
MSAT	macroscopics slide agglutination test
MUM	maximum unique matches
MUMi	maximum unique matches index
MW	molecular weight
ng	nanogram(s)
NGS	next generation sequencing
NO	nitric oxide
OD	optic density
OIE	World organisation for animal health
ORF	open reading frame
PAMP	pathogen associated molecular patterns
PCR	polymerase chain reaction
PFGE	pulse field gel electrophoresis
PG	prostaglandin
pIII	M13 minor coat protein III
PRR	pattern recognition receptors
PT	prothrombin time
pVIII	M13 major coat protein VIII

RAPD	random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RND	resistance nodulation division
RPKM	reads per kilo base per million mapped reads
rpm	rounds per minute
rRNA	ribosomal RNA
RT	room temperature
s	second
SDS	sodium dodecyl sulfate
SEREX	serologic expression cloning
SNP	single nucleotide polymorphism
T3SS	type 3 secretion system
T4SS	type 4 secretion system
T5SS	type 5 secretion system
TLR	toll-like receptor
TNF	tumor necrosis factor
TPM	transcripts per million
U	unit(s)
UV	ultra-violet
V	volt(s)
v/v	volume per volume
VFDB	virulence factor database
VNTR	variable number of tandem repeats
WHO	World Health Organization
w/v	weight per volume
Δ	delta, deletion
λ	phage lambda, wave length
µg	microgram(s)
µL	microliter(s)
µM	micromolar
%	percent
°C	degree celsius
2D-PAGE	2 dimentional polyacrylamide gel electrophoresis

List of Tables

List of Figures

1.0 INTRODUCTION

1.1 Overview of *Leptospira spp.* and Leptospirosis

1.1.1 Epidemiology

Leptospirosis is the most common zoonotic disease worldwide. Humans get infected by exposure to *Leptospira spp*. through contact to body fluids of carrier animals and contaminated water and soils (Mwachui et al. 2015). Pathogenic *Leptospira* spp. colonise the proximal renal tubules of reservoir hosts and are excreted through urine into the external environment, in which they can survive in water for several months (Trueba et al., 2004 and Andre-Fontaine et al. 2015).

Global incidence of leptospirosis is estimated to be approximately 1.03 million cases per year, with a mortality of 5.7% (Costa et al., 2015). In Malaysia, cases of human leptospirosis have been reported since 1928 (Fletcher et al., 1928). The annual incidence of leptospirosis in Malaysia ranged between 1 and 10 per 100,000 population (Lim, 2011). As leptospirosis is one of the most prevalent infectious diseases in Malaysia, it was classified as a notifiable disease in 2011. In 2015, 5,370 cases and 30 human deaths due to leptospirosis were reported, although there are still some undiagnosed cases, especially in the mild form of the disease due to the nonspecific symptoms of mild disease (Garba et al., 2017).

Human leptospirosis predominantly occurs through the contact of abraded skin or mucosal membranes or conjunctiva with water or moist soil that is contaminated with the urine of carrier and infected animals (Ko et al., 2009). Few factors contribute to the maintenance and transmission of leptospirosis. A wide range of reservoir and carrier hosts play a significant role in maintaining the disease in a community. Rodents and non-rodents carry leptospires in their proximal renal tubule, and excrete a large number of live leptospires in their urine and contaminate the soil and water of their nearby habitat (Perez et al., 2011).

Climate change such as heavy rain and flooding is another factor that causes an increase in outbreaks as well as sporadic cases. These problems are normally occurring in tropical countries. Congestion to the drainage may exacerbate the condition and cause areas with stagnant water contaminated by urine from carrier

hosts (Reis et al., 2008). Contact of humans with contaminated water and soil or animal fluids may also be possible through occupational, recreational and sports activities.

Urban outbreaks of severe leptospirosis occur annually in Brazil and are mainly associated with intense periods of high rainfall and poor living conditions (Hagan et al., 2016). Tourists from non-endemic countries travelling to endemic countries might be highly predisposed to leptospirosis as they have not been exposed to bacteria previously (naïve immunities) (Ricaldi et al., 2006).

1.1.2 Morphology

Leptospira spp. belong to the phylum Spirochaetes along with *Borrelia burgdorferi* (the causative agent of Lyme disease) and *Treponema pallidum* (the causative agent of syphilis). The hallmarks of the phylum are flat wave- spiral shape and endoflagellar motility. The genus *Leptospira* comprises 22 species that are divided into more than 300 serovars (Cameron, 2015).

Leptospira spp. are thin, right-handed helical, motile organisms that have a diameter of 0.15 µm and a length of 10–20 µm. The cell envelope is similar to Gram-negative bacilli in which the inner membrane and peptidoglycan cell wall are closely associated and overlaid by an outer membrane that contains surface-exposed lipoproteins and lipopolysaccharide (LPS) (Ko et. al, 2009 and Haake et al, 2010).

The peptidoglycan layer is mostly composed of the disaccharide tri-peptide *N*-acetylglucosamine (GlcNAc)–*N*-acetylmuramic acid (MurNAc)–l-Ala–d-Glu–*meso*-diaminopimelic acid and the crosslinked dimer GlcNAc–MurNAc–l-Ala-d-Glu–*meso*-diaminopimelic acid–d-Ala–*meso*-diaminopimelic acid–d-Glu–l-Ala–MurNAc–GlcNAc (Slamti et. al, 2011, Cameron, 2015). Peptidoglycan macromolecule in *Leptospira* spp. is intrinsic for determining their helical morphology and retaining the helical shape of intact cells. In addition, *MreB* filaments are essential for the division of, and for the maintenance of cell shape in most bacteria including *Leptospira* spp (Slamti et. al, 2011).

The endoflagellum is a unique characteristic of the phylum Spirochaetes. It is composed of three main substructures: the basal body, a flexible hook and a

filament. The flagellar motor is larger and more complex in spirochetes while the basal body core components are conserved in Gram-negative bacilli. The basal body also has a periplasmic structure i.e. a collar anchored to the inner membrane and MS ring (Zhao et al., 2014)· The hypothesis of the function is for orientation of endoflagella in the periplasmic space (Moon et al., 2016).

Lipopolysaccharide (LPS) is a major component of the leptospiral outer membrane (OM), a polysaccharide layer that lies at the surface of the OM. LPS mediates agglutination with LPS-specific antibodies or antisera and is the basis for serotyping with the microscopic agglutination test (MAT) and serological diagnosis. On the other hand, LPS may get altered by simply adding LPS-antiserum to a leptospiral culture, inducing growth of escape mutants (Adler, 2015; Cameron, 2015). The expression of intact LPS was concluded to be essential for leptospiral survival *in vitro* and *in vivo* based on findings that most LPS mutants are nonviable for growth in culture (Murray et al., 2010 and Adler, 2015).

As in Gram-negative bacteria, leptospiral LPS consists of three components: lipid A, the core, and the polysaccharide. The key difference of leptospiral lipid A to that of the typical one found in other bacteria is that *L. interrogans* converts the usual GlcNAc (N-acetylglucosamine) disaccharide backbone of lipid A to GlcNAc3 so that each of two sugars has two amino groups and four amide-linked fatty acids. This is has been observed in environmental bacteria. Furthermore, the length and number phosphate residues of leptospiral lipid A differ from that of Gram-negative bacteria.

Lipoproteins are another component of proteins in leptospiral outer membranes. They are proteins that have been post-translationally modified by fatty acids at cysteine residues. The fatty acid residues of lipoproteins are hydrophobic and embedded into membrane lipid bilayers as an anchor associated with the membrane, while the protein components project out from the membrane (Matsunaga et al., 2002; Adler, 2005). The most abundant lipoprotein in pathogenic *Leptospira spp.* is LipL32, which serves as a large calcium sink for leptopiral cells (Adler, 2011). It is expressed during infection (Haake et al., 2000) and is recognized to be one of the most dominant seroreactive antigens during acute and convalescent leptospirosis (Lessa-Aquino et al., 2013).

1.1.3 *Clinical Manifestations & Complications*

Leptospirosis ranges in severity from mild, self-limited febrile illness to a fulminant life-threatening illness. The disease may involve a broad array of organ systems, reflecting the systemic nature of the infection. The incubation period from exposure to onset of symptoms varies from seven to twelve days. It can also be as short as three days or as long as thirty days (Haake and Levett, 2015).

The pathogenesis started by pathogen entry by penetration of tissue barriers to gain entrance into the host via a cut or abrasion and the mucous membrane of conjunctivae or oral cavity or swallowed water (Corwin et al., 1990; Lingappa et al., 2004; Stern et al., 2010). The first stage is when leptospires can be found in the bloodstream, which can last up to 15 days (Agampodi et al., 2012; Bharti et al., 2003). The second stage of the disease is the immune phase when the antibodies can be detected and leptospires start to be cleared in the bloodstream, also known as convalescent stage (Levett, 2001). This normally occurs around the second week after the onset of symptoms. IgM antibodies are the first to rise before the IgG and usually persist for several months (Silva et al., 1995).

The first phase of leptospirosis coincides with the period of leptospiremia. At this phase patients may present will nonspecific symptoms which are found in other general tropical diseases, i.e. an acute febrile illness characterized by fever, chills, and headache. It is also accompanied by cough, muscle ache (particularly lower leg), gastrointestinal symptoms (nausea, vomiting, diarrhea, abdominal pain), conjunctival injection (also referred to as conjunctival suffusions), jaundice and rash. As these symptoms are generally present in other tropical illnesses, leptospirosis is often confused with illnesses such as dengue, typhoid, malaria etc. (Haake and Levett, 2015).

Leptospirosis patients usually present with mild to moderate elevations in liver transaminases and direct (conjugated bilirubin). The frequency of jaundice varies among cases, most likely due to virulence properties of the specific *Leptospira* spp. Katz et al., 2001 reported that serogroup Icterohaemorrhagiae has a strong association with jaundice and bilirubin elevation. Jaundice also occurs due to acute

haemolytic anemia, in particular in patients who also have glucose-6-phosphate dehydrogenoase deficiency. Many patients present with leukocytosis and thrombocytopenia. Leukopenia in the setting of thrombocytopenia and anemia are suggestive of bone marrow suppression. Kidney functions are usually affected by raised serum blood urea nitrogen and creatinine levels. Urinalysis findings may include detection of pyuria, hematuria and elevated urine protein levels (Katz et al., 2001).

Although some patients develop rapidly progressive severe disease with a high case fatality rate, other infections can persist in a latent asymptomatic state, which suggests that leptospirosis may also develop a chronic infection in humans (Adler et al., 2012). Due to the variable presentations, the World Health Organization (WHO) has classified leptospirosis into four broad clinical categories based on its severity: (1) a mild, influenza-like illness (ILI); (2) Weil's syndrome characterized by jaundice, renal failure, haemorrhage and myocarditis with arrhythmias; (3) meningitis/meningoencephalitis; (4) pulmonary haemorrhage with respiratory failure (WHO, 2003).

Weil’s syndrome is a combination of jaundice, acute kidney injury, myocarditis and haemorrhage symptoms. Severe headache accompanied by meningismus is suggestive of leptospiral meningitis/meningoencephalitis. Depending on the epidemiological setting, leptospirosis may be the cause of aseptic meninigitis (Silva 2002a). Patients with altered mental status may be affected by meningoencephalitis. Other neurological complications include hemiplegia, transverse myelitis, and the Guillain-Barre syndrome (Levett et al., 2001).

Severe leptospirosis may leads to circulatory collapse (septic shock) accompanied by acute respiratory distress syndrome (ARDS). Severe pulmonary haemorrhage syndrome due to leptospirosis may feature fatality rates of ≥50% (Gouveia et al., 2008). Chest radiographs may show diffuse alveolar infiltrates. Pulmanary haemorrhage may require lung drainage, endotracheal intubated and mechanical ventilation.

1.1.4 *Genomic and molecular aspects of pathogenesis*

The earliest *Leptospira* genome to be sequenced were *L. interrogans* serovar Lai (Ren et al., 2003) and Copenhageni (Nascimento e al., 2004). Both belong to the serogroup of Icterohaemorrhagiae and exhibit 95% identity at nucleotide level. It consists of a large circular chromosome (4,277 kb, 35 mol% GC) and a smaller replicon (350 kb, 35% GC). The majority of *L. interrogans* chromosomes are collinear except for a few gaps and a large inversion in the large chromosome (Nascimento et al., 2004). Later, more complete genome sequences are available including *L. biflexia* (Picardeau et al., 2008), the intermediate *L. licerasiae* (Ricaldi et al., 2012a) and other pathogenic species of Leptospira (Bulach et al., 2006; Chou et al., 2012) that provide more insights into the three different phynotypes of *Leptospira spp.* i.e. pathogenic, intermediate and saprophytic.

Leptospira spp. have genome of >3.9 to 4.6Mb characterized by 35-42 mol% G+C content which have a gene density of 75-92% of housekeeping function. The small second replicon (cII) range size from 278 to 350 kb and carries essential genes such as metF (methylene tetrahydrofolate) and asd (aspartate semialdehyde dehydrogenase) (Bourhy and Saint Girons, 2000; Zuerner et al., 1993). A third circular replicon, called p74 (74 kb, 36% mol% GC) has been identified only in *L. biflexia.* These replicons carry core genes located on the large chromosomes in the *Leptospira* species and they have nucleotide composition and codon usage that are very similar to those of large chromosome with which they are associated.

In contrast to other bacteria, 16S (*rrs*), 23S (*rrl*) and 5S (*rrf*) rRNA genes are clustered and co-transcribed, those in *Leptospira* are widely scattered on the large chromosome (Baril et al., 1992a). The slow-growing pathogenic and faster growing intermediates and saprophytic species of *Leptospira spp*. have a similar number of transfer RNA (tRNA) (Picardeau et al., 2008; Ricaldi et al., 2012a). Several insertion sequences such as IS1500, IS1502 and IS1533 have been identified In *Leptospira* (Boursaux-Eude et al., 1995; Zuerner and Huang 2002; Zuerner 1994). The number of insertion sequences in *L. borgpetersenii* is much higher than those in L. interrogans and *L. biflexia* which may be the answer to the genome deletions or rearrangement in *L. borgpetersenii* (Bulach et al., 2006). There are 3718 coding

sequences (CDS) being detected in serovar Lai strain 56601 (Ren et al., 2003; Zhong et al., 2011). A high-passage of virulence – attenuated isolate of serovar Lai revealed that there were insertions, deletions and single nucleotide variations detected in 101 genes (2.7% of total gene content (Zhong et al., 2011). DNA microarray hybridization revealed that there is high similarity in gene content among 11 *L. interrogans* strains from different serovars (He et al., 2007). Gene redundancy and gene duplication are also seen in *L. interrogans* (Picardeau et al., 2008).

The genome sequences of *L. biflexia, L. interrogans, L. borgpetersenii* contain approximately ~40% genes of unknown function. function In fact of 627 genes unique to *L. interrogans*, more than 500 (80%) encode hypothetical function is unknown in the genes unique to pathogenic proteins. This is consistent with the view that Leptospira possesses unique virulence factors which cannot be identified by similarity to those of other bacteria (Adler and de la Pena Moctezuma, 2010)

The molecular basis of leptospirosis pathogenesis is still largely unsolved. Due to the huge phylogenetic distance with other well studied bacterial pathogens, and the fact that 40% of its whole genome correspond to unknown elements or genes encoding hypothetical proteins, the uncovering on how disease are caused by pathogenic *Leptospira spp.* is still a huge challenge.

Due to the advent of genetics, studies on mutants and mutagenesis have been conducted and the characterization of virulence factors of pathogenic Leptospira spp, has been initiated (Bourhy et al., 2005; Croda et al., 2008). 13 virulence factors have been identified through these methods, by the principle that mutation of these factors has led to attenuation of virulence in the pathogen. These factors are; Loa22, HemO, FliY, la1641, lman1408, ClpB. FlaA2, KatE, Mce, LruA, HtpG, ColA, and lb194. These factors play different roles as predicted virulence factors, and their pathogenetic mechanisms still remain to be defined.

Studies on virulence factors of *Leptospira spp*., have been carried out by bioinformatics prediction by comparing pathogenic and saprophytic genomic make-up as well as experimental identification and confirmation. Genomes of pathogenic leptospires were noted to also encode spingomyelinases, phospholipases, other

proteases, TlyABC-like hemolysins as well as unusually large number of leucine-rich repeat proteins (Nascimento et al., 2004). Being pathogens, it is still a mystery how leptospires translocate effector molecules into host cells as there is lack of recognition systems of secretion system type III, IV and VI (Nascimento et al., 2004).

The majority of putative virulence factors that was identified do not have a role in the host. The readout of mutagenesis and in vivo testing for virulence factors with this property is that their inactivation does not lead to attenuation or improved animal survival (Adler et al., 2011). This has led to the theory that these virulence factors most likely work in functional redundancy, which may have resulted from genomic expansion through gene duplication (Bulach et al., 2006). For instance, LipL32 and LigB are obvious virulence factors by way of in vitro functional characterization, conservation and expression profiles. However, the mutants in these genes retain their full virulence in both acute disease and animal colonization models (Croda et al, 2008; Murray et al., 2009c). To add to this complexity, virulence factors that prove positive in one serovar, may not be functionally important in another serovar. For instance, Lourdault et al., 2001 reported that the clpB mutant is avirulent in serovar Kito but retains virulence in serovar Manilae.

Motility is a very important component of leptospires that is associated with its virulence. Leptosprires are highly motile due to presence of two periplasmic flagella. Loss of flagellar function results in loss of motility and attenuation of virulence, as demonstrated by mutant studies of various motility genes i.e. flaB, fliY, and flaA2 (Picardeau et al., 2001). Leptospires possess the majority of key chemotactic genes found in other bacteria (Nascimento et al., 2004; Ren et al., 2003). Leptospires respond to a variety of clinical stimuli, for instance hemoglobin which results in attraction to site where tissue barriers have been degraded during pathogen entry (Yuri et al.,1993; Lambert et al., 2012b). It is possible that chemotaxis is not required when host has already been invaded, as mutants of the cheB and cheX genes did not attenuate virulence in hamsters infected intraperiotneally (Murray et al., 2009a). However, this could also be due to the presence of multiple cheB genes encoded in the leptospiral genome.

Adhesion of leptospires to host surfaces is an important step for its pathogenesis and correlates positively with strain virulence (Tsuchimoto et al.,1984). Leptospires can adhere to various components of extracellular matrix (ECM) including laminin, fibronectin, proteoglycans and 28 types of collagen, with Type I, III, IV and IV being most prominent (Batzios et al.,2013). There are groups of leptospiral proteins with multiple binding affinities which include Lig proteins, Len proteins and LipL32. Besides binding to host ECM tissues, Lig proteins also bind complement regulatory and play role in tissue damage by binding to the fibrinogen and matrix components associated with wound healing (Castiblanco-Valencia et al., 2012; Choy, 2012).

1.1.5 *Typing*

Identification and typing of *Leptospira* species play an important role in understanding disease epidemiology and pathogenicity, together with the development of diagnostic tools, effective vaccines and preventive strategies. Knowledge of the prevalent serovars and their maintenance hosts is essential for understanding the epidemiology of leptospirosis (Romero et al., 2009). This knowledge can lead to the recognition of carrier mammals and enable targeted prevention methods to contain outbreaks and is important in identifying new species and serovars (Galloway and Levett, 2010).

In the past three decades many molecular typing methods have been developed for *Leptospira spp.* typing. These include DNA-DNA hybridization (DDH) analysis (Yasuda, et al., 1987; Ramadass et al., 1992; Brenner et al., 1999; Perolat et al., 1998), randomly amplified polymorphic DNA (RAPD) fingerprinting (Corney et al., 1993), arbitrarily primed PCR (AP-PCR) (Ralph et al., 1993; Perolat et al.,1994), pulsed field gel electrophoresis (PFGE) (Hermann et al., 1991), restriction fragment length polymorphism (RFLP) analysis (Zuerner et al.,1990), bacterial typing methods based on insertion sequences (IS) (Zuerner et al., 1995), detection of variable number of tandem repeats (VNTR) (Majed et al., 2005; Slack et al., 2006), *rrs* sequencing (Hookey et al., 1993, Morey et al., 2006; Cerquiera et al., 2010) and sequencing of specific genes or gene fragments including *rpoB*, *gyrB*, *secY* and *ligB* (La et al., 2006; Slack et al., 2006; Victoria et al., 2008; Ahmed et al., 2009; Cerquiera et al., 2009; Ahmed, et al., 2011).

The genus *Leptospira* was first described and classified based on morphological observations (Picardeau, 2014). Phenotypically, the *Leptospira* genus was initially divided into pathogenic and non-pathogenic species, and later further divided into more than 300 serovars, based on the structural heterogeneity in the carbohydrate component of their lipopolysaccharide (LPS) (Levett 2001; Bharti et al., 2003). Generally, the pathogenic *Leptospira* serovars have specific host preferences (for example, serovar Hardjo and cattle, serovar Canicola and dogs, and serovar Icterohaemorrhagiae and rats). However, these associations are not absolute and the molecular basis for such host specificity is unknown (Picardeau, 2017).

The determination of a serovar is based on the cross-agglutinin absorption test (CAAT). This method is too laborious and time-consuming to be used for routine typing. Moreover, antigenically related serovars (same serogroup) may actually belong to different species as a result of the horizontal transfer of the LPS biosynthetic locus (*rfb*) (Llanes et al., 2016). The application of pulsed-field gel electrophoresis (PFGE) showed that macrorestriction profiles largely coincide with specific serovars (Galloway et al., 2008; Hermann et al., 1992).

DDH and 16S rRNA phylogenetic analyses have divided the genus *Leptospira* into three distinct clades comprising 22 species (see the figure). The clades were determined from a maximum-likelihood phylogenetic tree of the *Leptospira* genus, which was based on the concatenation of a selection of 491 core genes. In one clade, there are 10 pathogens that can infect and cause disease in humans and animals; this pathogenic *Leptospira* spp. can be further divided into four subgroups (subgroups I–IV) (Xu et al., 2016). In another clade, there are five intermediates that have been isolated from humans and animals that may cause various mild clinical manifestations of leptospirosis. In the third clade, there are seven saprophytes *(L. idonii*) for which the genome sequence is not available (indicated in the figure) and which are unable to cause disease (Schmid et al., 1986; Paster et al., 1991; Yasuda et al., 1987; Brenner et al., 1999).

Multilocus sequence typing (MLST) shows phylogenies that are consistent with current species designations and provides strain resolution in a given species (Ahmed et. al, 2006). Matrix-assisted laser desorption/ionization–time-of-flight (MALDI–TOF) mass spectrometry can also be used to define *Leptospira* species

(Rettinger, et. al, 2012). Next-generation sequencing (NGS) will soon overtake phenotypic methods for the identification and characterization of bacterial isolates (Richter et. al, 2009). Indeed, *L. mayottensis* was identified as a new species following NGS and subsequent *in silico* analysis, which revealed genomic similarity values (genome-to-genome distances from draft genomes) instead of the classical method of DDH (Bourhy et. al, 2009). With the availability of an increasing number of genome sequences for all leptospiral serovars, we can soon expect to decipher the genetic basis for LPS antigenic diversity, thus enabling the development of molecular methods to identify *Leptospira* strains at the serovar level.

Due to the advent of whole-genome sequencing-based methods, laborious and time-consuming methods such as DDH and CAAT will soon be substituted with *in silico* methods using whole genome sequences data. In fact an *in silico* DNA-DNA hybridization method has been published and established (Auch et al., 2010). Furthermore, by whole genome sequence as a gold standard of typing, less time and cost will be spent with accurate results and efficient and expedite the time for typing of new species and serovars.

There are a variety of approaches to analysing these genomic data for epidemiologic and infection control purposes. At a high level, phylotyping (strain interrelationships based on sequence- associated evolutionary history) may support epidemiologic investigations to determine the source and routes of infections, trace cross-transmission of healthcare-associated pathogens and identify virulent antibiotic-resistant lineages or subpopulations

1.16 *Diagnostic tools*

The need for improved diagnostic tests has been the subject of great awareness when multiple outbreaks were reported, especially in South East Asia (Agampodi et al., 2011; Amilasan et al., 2012). The outbreaks are usually linked with natural disasters such as floods and hurricanes (Lau et al., 2010; Levett, 2001). Due to its wide range of clinical presentations, the features of leptospirosis usually overlap with other common tropical febrile illnesses, i.e. dengue, chikungunya, malaria, and rickettsioses. In most of the case, clinical findings alone are not enough to reliably predict the pathogen based on clinical signs and symptoms (WHO, 2003).

For the clinicians, accurate and easy to use laboratory diagnostics tools are the answer to overcome missed or under-diagnostic cases, as these constitute an obstacle to the understanding of the natural history of the infection. Furthermore, an early and accurate diagnosis will assist clinicians to start the right antibiotics without delay. This will not only contribute to better patient outcomes, but also minimizes the overall cost of hospitalization (Bharti et al., 2003). For the public health authorities, accurate and efficient case detections will bring better classifications of case definitions and ultimately control strategies and case management, particularly in epidemic situations.

The gold standard of most bacteria detection, i.e. culture, cannot be used in diagnostic settings in the case of leptospirosis. This is due to its slow-growing nature and the long incubation time required *in vitro*, which may range from 1 month to 3 months. On the other hand, microscopic agglutination test (MAT), the gold standard method for serology detection requires time for anti-leptospira antibodies develop in the patient, which is usually in the second week after onset of symptoms (Levett, 2001). Moreover, the MAT requires specific laboratory equipment, a series of endemic serovar strains as well as highly trained staff. Usually, this service is offered only in a reference laboratory for leptospirosis.

The choice of diagnostic tests mostly relies on the performance of the diagnostic test itself, duration of time for the results to be produced, the practicality of the test depending on the type of sample that is available, and the cost involved.

Serology-based rapid diagnostic tests such as lateral flow immunochromatography, enzyme-linked immune-sorbent assays (ELISAs), and latex agglutination are the most common and popular tests at peripheral healthcare centres. The advantage of rapid tests is that they are rapid and simple and can be used for screening of suspected cases. However, they usually lack specificity and positive cases therefore need to be followed up by confirmation tests with high specificity.

The disadvantage of serological tests has been reported with varying diagnostic performances in distinct countries or regions. The reasons for this can be due to population-related differences such as past exposure to leptospirosis or other infectious agents, person-to-person bias, and insufficiently trained personal

(Picardeau et. al, 2014).

For whole-bacteria–derived antigens; false positives results may be due to persistent or cross-reacting antibodies. While this problem can be recovered with a conserved recombinant antigen, on the other hand, it will enable more specific test results with antigen-homologous infections. Variants of serovars among *Leptospira species* will reduce the sensitivity of the test. A validation from a local perspective prior to introduction of a new test should be performed using serum banks of the specific endemic region. This is to ensure that the test can be useful for the serovars variants in the region (Picardeau et. al, 2014).

The diagnostic tools can also be classified based on the two phases of leptospirosis i.e. leptospiraemia and antibody phase. Leptospiremia phase is the time between infection and the point antibody start to rise. During this phase, leptospires can be found circulating in the blood of the patient. In theory, leptospirosis can be diagnosed by dark-field microscopy of blood taken during the first week of illness. The limit of detection was determined as approximately 10^4 leptospires / mL of blood or urine. Unfortunately, it is practically unreliable as Brownian movement of collagen fibrils and other artefacts can resemble viable leptospires (Vijayachari et al., 2001). PCR-based methods are becoming widely used for early diagnosis of leptospirosis in the acute phase, due to their high sensitivity and specificity. Conventional and real-time PCR leptospirosis diagnosis assays have been developed, targeting whole gene arrays and conserved pathogenic genes such as *secY* and *lipL32* (Ahmed et al., 2012; Ahmed et al.; Slack et al., 2007; Thaipadungpanit et al., 2011; Villumsen et al., 2012). The limit of detection of PCR assays was generally determined as 100–1000 bacteria per ml of blood or urine (Bourhy et al., 2011; Slack et al., 2006; Smythe et al., 2002; Stoddard et al., 2009). Quantitative leptospiremia (bacterial load) can also be used for diagnosis. However, it does not correlate with clinical prognosis (Agampodi et al., 2012; Segura et al., 2005; Truccolo et al., 2001). A positive PCR usually indicates that pathogenic *Leptospira* is present in the sample but a negative PCR does not rule out leptospirosis. Identification of leptospiral genotype or serovar can be performed molecularly using melting curve analysis (Merien et al., 2005) and DNA sequencing (Perez and Goarant, 2010). Isothermal amplification techniques have been applied to leptospirosis diagnosis as well (Colenbrander et al., 1994;

Koizumi et al., 2012; Lin et al., 2009; Sonthayanon et al., 2011). Due to the advantage of being an alternative to PCR-based methods using a single constant temperature, this method is suitable to be used in resource limited centers (Mori and Notomi, 2009).

The second phase is when, the detection of antibody has been the method of diagnosis based on the presence of specific antibody. The in-house IgM ELISA based on a whole cell antigen extract has been the reference screening test in most leptospirosis reference centers (Goris et al., 2012). Normally, the antigens that are used in the test are recombinant surface proteins or lipoproteins of *Leptospira* spp and generally the sensitivity of these tests is poor (Levett, 2001; McBride et al., 2005). A serological diagnosis of leptospirosis made by a screening ELISA needs confirmation through microagglutination test (MAT), polymerase chain reaction (PCR), culture or by paired samples of both acute and convalescence phase. Despite the inconsistent results of ELISA, studies have found that the test can detect antibodies much earlier than MAT (Aviat et al., 2010; Bajani et al., 2003; Cumberland et al., 1999; Doungchawee et al., 2008; Levett, 2001; Signorini et al., 2013). Anti-*Leptospira* IgM also can be detected earlier than IgG and agglutinating antibodies (Silva et al., 1995). Other rapid antibody detection methods include immunofluorescence assay (IFA), macroscopic slide agglutination test (MSAT), indirect hemagglutination assay (IHA), microcapsule agglutination test (MCAT), complement fixation test (CFT), and lateral flow assays (LFA). IFA use antibodies conjugated to a fluorescent dye as a detection reagent to detect specific antibodies in body fluids or antigen in tissue sections. The sensitivity and specificity of the indirect fluorescent antibody (IFA) test correspond to those of the ELISA (Appassakij et al., 1995). MSAT uses dense suspensions of killed *Leptospira spp.* serovars to detect antibodies in patient serum (Wolff and Bohlander, 1966). It is relatively insensitive for diagnosis but may be useful for epidemiological screening (Marin-Leon et al., 1997). IHA uses red blood cells sensitized with an extract of an erythrocyte-sensitizing substance from *L. biflexa* serovar Patoc and detects antibodies in heat-inactivated serum by the presence of agglutination (Chang et al., 1957; McComb et al., 1957; McBride et al., 2007). MCAT uses a synthetic polymer coated with a mixture of the whole cell-derived antigens from pools of serovars. The test showed high sensitivity, however failed to detect infections by certain serovars

(Arimitsu et al., 1982; Arimitsu et al., 1994). CFT can be used for antibody and antigen detection. The test is based on the presence of antigen-antibody complexes, which leads to reaction with the complement system and prevents cell lysis. However, it is laborious and lacks diagnostic accuracy (Andreescu, 1990). LFA is one of the most commonly used rapid test as it can be performed at the patient bedside. Antigen based on whole cell antigen or immunogenic antigen enables the rapid detection of *Leptospira*-specific IgM antibodies in human sera (Smits et al., 2001), but variable diagnostic performance has been reported (Goris et al., 2013).

1.2 ORFeome Phage Display

1.2.1 Introduction

Open reading frame (ORFeome) phage display is a phage display approach to identify immunogenic epitopes in open reading frames. The fundamental technology was introduced in 1985 for filamentous M13 phage (Smith et al., 1985). It employs non-lytic filamentous M13 phage for fusion of heterologous protein to a phage coat protein, mostly minor coat protein III (pIII). The phage particles contain the genotypic information of the oligopeptides which are displayed on the surface of the phage as the phenotype (Breitling et al., 1991). This approach can be used to display oligopeptides and/or full protein domains from genomic DNA or cDNA.

Firstly, DNA of interest is fragmented randomly, resulting in a random directional manner of DNA fragments. After this, only about one out of eighteen clones results in an open reading frame (ORF). Thus, enrichment of ORFs is a vital step in order to improve library quality. Several different strategies can be taken to accomplish ORF enrichment, such as cloning the DNA fragments upstream of a selection marker like β-lactamase. The expression of functional β-lactamase confers ampicillin resistance, and correct folding allows the corresponding E. coli clone to grow on selective media. , are the signatures of fragments cloned in-frame with the β-lactamase gene. Enriched ORFs are later subcloned into a phage display vector (Faix et al., 2004; D'Angelo et al., 2011) or the β-lactamase gene is removed by CRE-mediated recombination (DiNiro et al. 2010). In a T7 phage display approach to identify Calpain substrates, fragmented DNA was cloned upstream of a biotinylation-tag. Furthermore, immobilizing produced phage on a surface using streptavidin allowed

an ORF selection and the identification of protease substrates by Calpain-mediated release from the surface. Infectivity of the lytic M13 phage is mediated by the pIII protein whereby each phage particle has up to five pIII copies. The N-terminal domain can be divided into two subdomains (N1 and N2) that mediate infection by binding to the host cell's F pilus, whereas the C-terminal domain (CT) anchors the protein in the phage capsid. The three regions are separated by glycine and serine rich linkers. Introducing trypsin sites into the linker regions renders pIII sensitive to proteolytic cleavage and decreases infectivity after trypsin treatment. Cloning randomly fragmented DNA upstream of a trypsin-resistant pIII gene in a phagemid vector and using a trypsin-sensitive helper phage for phage packaging allows ORF enrichment of genomic libraries after trypsin treatment and infection of *E. coli* host cells (Gupta et al., 2013).

Hyperphage is a special helpher phage, as it is infective and has pIII in its capsid as well as genomic deletion of the pIII encoding gene gIII "Hyperphage" (Rondot et al., 2001; Soltes et al., 2007), ORFs can be enriched without further cloning steps or protease treatment (Kügler et al., 2013; Hust et al., 2006).

Thus, fragmented DNA which were randomly positioned is cloned upstream of the pIII gene on a pHORF3 vector. Subsequent to "Hyperphage" co-infection, the infective phage particles are only being produced if there is an open reading frame, resulted by an in-framed DNA insert sequence with the pIII gene, and there was no stop codon in between. This is because, the fusion protein is the only pIII source to render the phage particles infective.

1.2.2 Advantages and limitations

This technology has been used successfully to identify novel potential biomarkers and a candidate vaccine against *Mycoplasma species* (Kügler et al., 2008; Nasseem et al., 2010), *Salmonella typhimurium* (Meyer et al., 2012), *Neisseria gonorrhoeae* (Connor et al., 2016), and from a tick salivary gland cDNA library (Becker et al., 2015). ORFeome phage display can be used for any ORF containing DNA study as it is independent of cultivation.

Additionally, ORFeome phage display may identify proteins with a molecular mass <10 kDa or weakly expressed proteins that are not be identified by 2D-PAGE/MS (Beranova-Giorgianni et al., 2003; Rabilloud et al., 2010; Urguhart et al., 1998).

The limitation of this method are oligopeptides have to be displayed on a phage in order to be displayed and identified. Generally, the epitope of peptides are being displayed on the surface of the phage and the full proteins are rarely displayed (Kügler et al., 2013; Hust et al., 2006).

1.2.3 Alternative Methods

The conventional studies of pathogenic proteins to be used for diagnotics applications are based on pathogen cultures followed by 2D-PAGE and immunoblotes analysis using patient sera nand mass spectrometry for protein identification (Delvecchio et al., 2006; Jacobsen et al., 2005; LaFrentz et al., 2011). Pathogen needs to be cultured in direct contact with the host in order to study and identify seroreactive /immunogenic antigens/proteins that are important to pathogenesis. This is to ensure the pattern of the gene expression reflects the real situation in an infection. The limitations of this method is that the pathogen proteome will most likely be overwhelmed by that of the host (Zhang et al., 2005). Additionally, proteins with a molecular mass <10 kDa or weakly expressed proteins may not be identified by 2D-PAGE/MS (Beranova-Giorgianni et al., 2003; Rabilloud et al., 2010; Urguhart et al., 1998) might not be identified.

As an alternative, microarrays and serologic expression cloning (SEREX) were two other methods of identification of immunogenic proteins from pathogens (Hoppe et al., 2013; Danckert et al., 2014; Kunnath-Velayudhan et al., 2010; Chen et al., 2005). Microarray screens are laborious and expensive as there were several hundreds of target recombinant proteins at a time. On the other hand, SEREX relies on the expression of cDNA libraries in E. coli using lytic phage vectors (Chen et al., 2005). The recombinant protein plaques are transferred to membranes and stained with patient sera. DNA sequencing of the corresponding clones were performed for identification of immunogenic protein. The drawbacks of this technology are handling of lytic phage, maintenance of the phage library, and laborious screening.

1.7 Objectives

1.7.1 Whole Genome Analysis of Leptospira spp.

General objective:

To perform deep molecular characterization of six *L. interrogans* clinical isolates in order to determine genotypic features that correlate with clinical phenotypes

Specific objectives:

i) To determine the most severe and mildest strain by clinical evaluations & animal study.

ii) To determine phylogenomic relatedness using typing based on whole genome sequences.

iii) To determine pathogenomic inferences of strains determined in (i) based on genomic and transcriptomic studies.

1.7.2 Biomarker Identification of Leptospira spp.

General objective:

To identify pathogen biomarker(s) of *Leptospira spp*. that are suitable to be used as a diagnostic tool in source limited, endemic countries.

Specific Objective:

i) To identify novel seroreactive peptides of *Leptospira spp.* using ORFeome Phage Display.

2.0 METHODS: WHOLE GENOME ANALYSIS OF *LEPTOSPIRA SPP.*

2.1 Clinical Cases and Laboratory Investigations

2.1.1 Sample collections

Approximately 2000 serum samples from hospitalized patients in Hospital Tuanku Ampuan Rahimah, Klang with positive IgM Leptospirosis ELISA (Panbio Pty., Ltd., Queensland, Australia) were cultured for *Leptospira spp*. isolation in Ellinghausen-McCullough-Johnson-Harris with 5-Fluorouracil (EMJH-5FU) media from 1st January 2014 to 31st December 2015.

Five isolates of *L. interrogans* were obtained from leptospirosis patients and inoculated with Modified EMJH media followed by an incubation at 30°C for four weeks as described before (Bejo et al., 2017). One isolate originating from Langkawi, Malaysia, was purchased from the Leptospirosis Reference Centre, Amsterdam (Wagenaar et al*.,* 2005).

2.1.2 Leptospira isolation

Two drops of patient serum to be sent for Microscopic Agglutination Test (MAT) for confirmation of leptospirosis were inoculated into 5 ml EMJH-5FU media and incubated in 30° C for four weeks.

All cultures were examined under a dark-field microscope for the presence of *Leptospira spp*. This method requires the presence of a minimum of 1 x 10^4 cells /mL for leptospires to be visible. This method requires trained personnel, otherwise the diagnosis can be missed (Bejo et al., 2017).

2.1.3 Culture media

The nutritional requirements of leptospires are simple and unique. Vitamin B1 and B12 and long chain fatty acids are the organic compounds known to be essential nutrients and phosphate, calcium, magnesium and iron are the essential inorganic components for the growth of leptospires. Since leptospires cannot synthesize fatty acids *de novo*, these compounds must be supplied to meet their carbon, cellular, lipid and energy requirement. However free fatty acids are inherently toxic to leptospires. Thus, these nutrients must be presented to the bacteria bound to

albumin. Polysorbate (Tween) can be included in the medium of leptospira cultures as the source of fatty acids. Carbohydrates are not suitable as a source of energy or carbon for leptospires (Bejo et al., 2017).

2.1.4 Culture media preparation

EMJH medium is prepared by mixing nine volumes of EMJH basal medium and 1 volume of albumin fatty acid supplement (AFAS). The EMJH basal medium was prepared by adding 996 mL of distilled water with 1.0 g Na_2HPO_4 (anhydrous), 0.3 g KH_2HPO_4 and 1 g NaCl. 1mL of each of the following stock solutions (NH4Cl, thiamine, sodium pyruvate, glycerol) was added and mixed well. The pH was adjusted to 7.5 ± 0.1 using 1M NaOH. 270 mL of the mixture was dispensed into a clean screw-capped bottle and autoclaved at 12°C for 20 minutes (min) and stored at 4°C until use.

Albumin fatty acid supplement (AFAS) was prepared by dissolving 20 g of bovine albumin fraction V powder (BSA) in 100 mL autoclaved double-distilled water. The following stock solutions were added i.e. 2 mL $ZnSO_4$, 2 mL $MgCl_2$, 2 mL $CaCl_2$, 2 mL $FeSo_4$, 20 mL $CuSO_4$, 2 mL Cyanocobalamin, 25 mL Tween 80. The mixture was mixed well and adjusted for pH 7.5 ± 0.1 using 1M NaOH. Distilled water was added to make up to 200 mL, and the solution was filtered using a 0.22 µM Millipore filter before storage. Sterility was checked by adding 1 mL of albumin supplement to the 10 mL nutrient broth at 37°C for 24 hours (h) before examining for bacterial contamination.

2.1.5 5-Fluorouracil stock preparation

1 g of 5 Fluorouracil (5-FU) was dissolved in 5 mL distilled water. 2 mL of 2N NaOH was added and heated gently to 56°C to facilitate dissolution. The pH was adjusted to 7.5 ± 0.1 with 1N NaCl, and the volume was added to 100 mL with distilled water. The solution was filtered using 0.45 µM Millipore filter and stored at 4°C.

2.1.6 EMJH-5FU liquid media preparation

To prepare the liquid medium of EMJH, 30 ml of AFAS was added to 270 mL of EMJH basal medium. 5 mL of the solution was dispensed into polypropylene conical

tubes or other appropriate sterile tubes. 5-FU was added to EMJH medium at a concentration of 200 to 400 µg/mL. Quality control was performed on 5% of the newly prepared medium by incubating the medium at 37°C for 3 days before examining for contamination. The medium was also inoculated with a pathogenic serovar to determine its ability to support growth.

2.1.7 Clinical data collection

The following clinical data were extracted from the patients' records:

Patients history:

Age, gender, occupation, history of travelling, jungle trekking, recreation at waterfalls or river.

Clinical history:

Presence and duration of fever, chills, headache, myalgia, arthralgia, conjunctival injection, anuria or oliguria and/or proteinuria, jaundice, pulmonary and/or intestinal hemorrhage, cardiac arrhythmia or failure, skin rash, and gastrointestinal symptoms such as nausea, vomiting, abdominal pain, and diarrhea

Physical findings:

General: temperature, blood pressure, pulse rate, respiratory rate, oxygen saturation, Glasgow coma scale.

Specific: myalgia (particularly associated with the calf muscles and lumbar region), arthralgia, conjunctival injection, meningeal irritation, anuria or oliguria and/or proteinuria, jaundice, hemorrhages (from the intestines and lungs), cardiac arrhythmia or failure, macular rash, epigastric tenderness.

Laboratory investigations:

White cell count, hemoglobin, prothrombin time, partial prothrombin time, amylase, albumin, total protein, alkaline phosphatase (ALP), alanine transferase (ALT), aspartate aminotransferase (AST), total bilirubin, sodium, potassium, urea, creatinine, creatinine kinase (CK), C-reactive protein (CRP), IgM leptospirosis test,

microagglutination test (MAT).

Radiological investigations:

Chest X-ray

Management:

Antibiotics administration, resuscitation with inotropes, ventilation.

2.2 Virulence Test on Animal Model

2.2.1 Ethical statement

The ethics for the animal experiments were authorized by 2nd Local Institutional Animal Care and Use Committee (JACUC), Kraków, Poland and conducted in compliance with Resolution NR 172/2017.

2.2.2 Preparation of leptospires for inoculation

All leptospires were cultivated in liquid media (EMJH) media (Difco Laboratories) incubated at 30°C in rocking incubator 200 rpm for 7-10 days to reach logarithmic phase. The concentration of the leptospires was determined by counting the number of leptospires using a Petroff-Hauser counting chamber under DFM as previously described (Faine, 1982). All six strains were prepared with a concentration of 10^8 leptospires/ml.

2.2.3 Leptospires Infection to Guinea Pigs

Three female Hartley guinea pigs, age around 3 weeks old and weight range 180 – 200 g from AnimaLab, Budapest, were assigned to six different groups. An additional group was created as a control group. Each group were housed in individual cages and fed standard guinea pig chow and water *ad libitum.* All six groups were inoculated with 500 µl of 10^8 /mL of six *L. interrogans* strain intra-peritoneally. The control group was inoculated with EMJH media alone.

Animals were monitored daily for signs of illness including weight loss and loss of mobility and are euthanized when they appear moribund. The guinea pigs were euthanized with sodium pentobarbital ≥100 mg/kg intraperitoneal routes and

subjected to post autopsy examinations. Kidneys, lungs, and liver were collected for polymerase chain reaction (PCR) analysis and fixed in 10% formalin for histopathological examination. Virulence of strains was determined by histopathological abnormalities, biochemistry and cytokines analysis

2.2.4 Biochemistry tests

Sera of the guinea pigs were collected via cardiac puncture for biochemistry and cytokines analysis. Sera were sent to the laboratory for the following tests: amylase, albumin, total protein, alkaline phosphatase (ALP), alanine transferase (ALT), aspartate aminotransferase (AST), lactate dehydrogenase (LDH), direct and indirect bilirubin, sodium, potassium, urea, creatinine, creatinine kinase (CK) and gamma-glutamyl transferase (GGT).

2.2.5 Real-time PCR

2.2.5.1 DNA extraction from tissues of guinea pigs

Tissue samples, i.e. liver, lung and kidney, that were stored in -80°C were thawed in a laminar flow hood. DNA was extracted from tissues using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). 180 µL buffer ATL was pipetted into a 1.5 mL microcentrifuge tube. A suitable sterile petri dish and sterile blade were used to cut up to 25 mg of tissue into small pieces. These small pieces were placed in a 1.5 mL microcentrifuge tube containing 180 µL buffer ATL. 20 µL proteinase K was added and mixed by pulse-vortexing for 15 s and incubated at 56°C in the heat block until the tissue was completely lysed. This took around 1-3 hours. Every 15 min the samples were vortexed to accelerate tissue lysis. Clear solution indicated complete lysis. The tubes were briefly centrifuged to remove drops from the lid of the tube. 200 µL Buffer AL was added and mixed by pulse-vortexing for 15 s and incubated at 70°C for 10 min. The tubes were centrifuged again to remove drops from the lid of the tube. Then, the genomic DNA extraction step similar to 2.3.1 was carried out.

2.3.5.2 Real-time PCR of Leptospira

Real-time PCR was performed on an iQTM5 Multicolour Real-Time PCR Detection System (Bio-Rad Laboratories, 2000 Alfred Nobel Drive, Hercules, CA94547 US) using the DNA-binding dye technique (SYBR Green). Reactions were performed in a

total volume of 25 µL consisting of 1x iQTM SYBR Green Supermix (Bio-Rad) of 2x stock reagent containing 100 mM KCl, 40 mM Tris-HCl, pH 8.4, 0.4 mM of each dNTP, 50 units/ml iTaq DNA polymerase, 6 mM MgCl2, 20 nM fluorescein and stabilizers. Forward and reverse primers as in (Ahmed et al., 2009) were added at a final concentration of 400 nM each. ICA was added in 0.5 µL volumes and DNA samples in 10 µL volumes. 10µL sterile water was used instead of DNA template as negative control. The amplification protocol consisted of 10 min at 95°C, followed by 40 cycles of amplification (95°C for 5 seconds (s), 54°C for 5 s, 72°C for 15 s). Subsequently, the reaction was stopped at 95°C for 2 min, cooled (20°C for 1 min) and melted (70–94°C with plate readings set at 0.5°C). The cut-off was set at cycle threshold (Ct) 38 that the last cycle without background noise. In order to determine the optimal concentration of the PCR reagents and the optimal annealing temperature, different reagent concentrations were tested in combination with a temperature gradient and various incubation times. All experiments were repeated at least twice. The duration of the final PCR cycles was approximately 100 min including generation of the melting curve. The resulting data were analyzed using the software provided by the iQ5 system (Bio-Rad iQTM5 2.0 Standard Edition Optical System Software, V2.0.148.060623).

2.2.6 Histopathology

Gross autopsy and histopathology examination of multi-organs (liver, spleen, kidney, lung, intestine) of all groups (including control) was performed. Infected and control groups of guinea pigs' viscera (kidney, lungs and liver) were fixed in neutral-buffered 4% formaldehyde, and processed routinely, embedded in paraffin and cut into 4 µm thick serial sections and stained with hematoxylin and eosin (H&E).

Paraffin-embedded tissue sections were stained with a modified Steiner silver stain for the presence of spirochetes (Nally et al., 2004). 10% formalin for histopathological examination.

2.2.7 Densiometric measurement

Five random 5x magnification views from pulmonary slides of all strains were captured and haemorrhage areas were measured using cellSens Standard software (Olympus Scientific Solutions Americas Corp). ANOVA was performed using

GraphPad Prism 7.0.

2.3 Whole Genome Sequence & Assembly

2.3.1 Genomic DNA extraction

30 mL of the densest possible 7-10 day culture of *L. interrogans* in EMJH medium was centrifuged. The pellet was suspended in 180 µL Buffer ATL (Qiagen, Germany). Genomic DNA extraction from isolates was performed using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany).

DNA extraction was performed in the same laminar flow hood in which the samples were thawed. If the amount of sample was less than 200 µL, sterile PBS was added to a final volume of 200 µL. 20 µL proteinase K (Qiagen Kit) was added and mixed by vortexing for 10 s. 200 µL Buffer AL was added and mixed by pulse-vortexing for 15 sec. The sample was incubated at 56°C in the heat block for at least 10-20 min. The sample was centrifuged briefly to remove drops from the inside of the lid. 200 µL ethanol (96-100%) was added and mixed by pulse-vortexing for 15 s and then briefly centrifuge. The mixture was carefully applied from step 8 to the QIAamp mini spin column, without wetting the rim, and centrifuged at 8000 rpm for 1 min. The spin column was placed in a clean 2 ml collection tube and the tube containing the filtrate was discarded. 500 µL buffer AW1 was added without wetting the rim and centrifuged at 8000 rpm for 1 min. The spin column was placed in a clean 2 mL collection tube and the tube containing the filtrate was discarded. 500 µL buffer AW2 was added without wetting the rim and centrifuged at 13000 rpm for 1 min. The spin column was placed into a 1.5 mL microcentrifuge tube from which the cap had been cut off and the tube containing the filtrate was discarded. The sample was centrifuged at 13000 rpm for 3-5 min to eliminate buffer AW2 carryover, making sure the spin column was dry. The spin column was placed into a new 1.5 mL microcentrifuge tube from which the cap had been cut off and the tube containing the filtrate was discarded. 100 µL DNAse free water was added and incubated at RT for 3 min then centrifuge for 1 min at 13000 rpm. The same collecting tube was used and 100 µL DNAse free water was added again, incubated at RT for 1 min, and the tube was then centrifuged for 1 min at 13000 rpm.

The DNA product was transferred from step 18 and 19 to a new 1.5 mL microcentrifuge tube with cap and briefly vortex and spin down and attach the label. The DNA is now ready to be used for the PCR or can be stored at -20°C.

2.3.2 PacBio library preparation and sequencing

SMRTbell™ template libraries were prepared according to the instructions from Pacific Biosciences (Menlo Park, CA, USA) following the Procedure & Checklist – Greater Than 10 kb Template Preparation. Briefly, for preparation of 15kb libraries 8µg genomic DNA was sheared using g-tubes™ from Covaris, Woburn, MA, USA according to the manufacturer´s instructions. DNA was end-repaired and ligated overnight to hairpin adapters applying components from the DNA/Polymerase Binding Kit P6 from Pacific Biosciences, Menlo Park, CA, USA. Reactions were carried out according to the manufacturer's instructions. BluePippin™ Size-Selection to greater than 4 kb was performed according to the manufacturer´s instructions (Sage Science, Beverly, MA, USA). Conditions for annealing of sequencing primers and binding of polymerase to purified SMRTbell™ template were assessed with the Calculator in RS Remote (Pacific Biosciences). Sequencing was carried out on the PacBio *RSII* (Pacific Biosciences) taking one 240 min movie for one-three SMRT cells per isolate using the P6 Chemistry.

2.3.3 Illumina library preparation and sequencing

For short reads, libraries for whole genome sequencing were prepared with the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® with 550 bp as insert size following the manufacturer's instructions. Sequencing using the MiSeq Personal Sequencer (Illumina Inc., San Diego, CA, USA) was done to 250 cycles in both directions. DNA-Seq reads were converted to FASTQ format and de novo assembled with the VELVET tool (Zerbino and Birney, 2008). The resulting contigs were annotated with the PROKKA software tool for annotation of prokaryotic genomes (Seemann, 2014).

2.3.4 Genome assembly, error correction, and annotation

SMRT Cell data was assembled using the "RS_HGAP_Assembly.3" protocol included in SMRT Portal version 2.3.0 using default parameters. The assemblies

revealed circular chromosomes and plasmids. All replicons were circularised, particularly artificial redundancies at the ends of the contigs were removed and adjusted to *dnaA* or the respective plasmid replication gene as the first gene. Error-correction was performed by a mapping of Illumina MiSeq data of 2 x 300 bp onto finished genomes using BWA (Li and Durbin, 2009) with subsequent variant and consensus calling using VarScan (Kobold et. al., 2012). A consensus concordance of QV60 could be confirmed for all of the genomes. Finally, annotation was carried out using PROKKA 1.8 (Seemann, 2014). The genome sequences will be deposited at NCBI.

2.3.5 RNA library preparation and sequencing

2.3.5.1 Total RNA extraction

4 mL of the densest possible 7-10 days culture of *L. interrogans* in EMJH medium was added to 8 ml of Qiagen RNAProtect (Qiagen, Hilden, Germany), incubated at room temperature for 5 min before centrifuge for 30 min ≥ 8000 x g. The supernatant was discarded and the pellet was stored at -80°C to preserve the RNA. Total RNA extraction from *L. interrogans* isolates was performed using the miRNAeasy Qiagen Kit (Qiagen, Hilden, Germany).

The culture pellet was suspended with 3 mL of TE Buffer and later centrifuged for 10 min at 8000 x g. The mixture was added with 150 µL TE Buffer Mixture (100 mL TE Buffer added with 20 µL Proteinase K (Qiagen, Hilden, Germany) and Lysozyme (15 mg/mL) to the pellet and resuspended. The mixture was incubated for 10min and vortexed for 10 sec every 2 min. 1 mL of Qiazol lysis reagent (preheated at 65°C) was added and the sample vortexed for 3 min and incubated for 5 min. 200 µL of chloroform was added and the sample was mixed rigorously by shaking. Then, the sample was incubated at room temperature (RT) for 3 min and centrifuged at 12,000 x g for 15 min at 4°C.

The upper, colourless layer of around 700 µL was transferred to a new tube and 1.5 volume of ethanol 100% was added. The mixture was vortexed. Around 700 µL of the sample (including any precipitate) was transferred to a spin column placed on 2 mL of a collection tube. The lid was closed gently and centrifuge for 15 sec at ≥ 8000

x g. The flow-through was discarded. This step was repeated using the remainder sample. 350 µL Buffer RWT (buffer have been added 45 mL Isopropanol) was added into a spin column and centrifuged for 15 sec > 8000 x g (to wash) and flow-through was discarded. DNAse I stock (added 550 uL RNase free water) was added with 70 µL Buffer RDD and gently mixed by inverting and not vortexed. 80 µL DNAse I was pipetted directly onto spin column membrane and it was incubated on the benchtop for 15 min. Then 500 µL Buffer RWT was pipetted into a spin column and centrifuged for 15 sec at ≥ 8000 x g. The flow-through was kept and reapplied to a spin column before it was centrifuged again at 15 sec ≥ 8000 x g. Then the flow through was discarded. 500 µL Buffer RPE was pipetted into the spin column and centrifuged for 15 sec ≥ 8000 x g, then the flow through was discarded.

Another 500 µL RPE was added into the spin column, centrifuged for 2 min >8000 x g to dry the column. The 2 mL collection tube was changed and centrifuged at full speed 1 min (to ensure the spin column membrane was totally dry). The spin column was transferred onto a new 1.5 mL Eppendorf tube. 30 µL RNAse free water was pipetted directly onto the column. The column was incubated for 10 min and centrifuged for 1 min >8000 x g for elution. The flow-through was pipetted onto the spin column and centrifuged for 1 min ≥ 8000 x g (to increase the RNA concentration).

2.3.5.2 RNA-seq library preparation and sequencing

The RNA Libraries of 300 bp prepared from total RNA using NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina® and Ribo-Zero rRNA Removal Kit by Illumina®. This protocols fragmentises the RNA, removes rRNAs, synthesises first and second strand cDNA and ligates adapters to the ends of the cDNA fragments. After QC on Bioanalyzer/Qbit templates go on the Illumina HiSeq2500 Sequencer system. Fragments are sequenced by 50 bp single end. The sequence output was controlled for general quality features using the fastq-mcf tool ea-utils (Aronesty, 2011) and was mapped against the genome sequence of the two strains *L. interrogans strains* using BWA v.0.7.17-r1188 and SAMtools (Li and Durbin, 2009) for storing nucleotide sequence alignments. Data were subsequently analysed with Rockhopper v.2.0.3 (McClure et al., 2013).

2.4 Typing & Phylogenetic Analysis

2.4.1 Serotyping

The isolates were typed at the WHO/FAO/OIE and National Collaborating Centre for Reference and Research on Leptospirosis (Amsterdam Medical Centre, Amsterdam, The Netherlands). To determine their presumptive serogroups, the *Leptospira* isolates were subjected to the MAT using a panel of rabbit anti-*Leptospira* sera that was representative of all pathogenic serogroups. The rabbit antisera (polyclonal) were prepared as described by (Terpstra et al.,1985; Faine et al.,1999). The MAT was conducted in a microtiter plate with serial 2-fold dilutions of the rabbit antisera, starting with a serum dilution of 1:10. Equal volumes of viable leptospiral strains and antiserum dilutions were mixed. After incubation at 30°C for 2 h, the mixtures were investigated for agglutination using dark-field microscopy. High rates of agglutination of the leptospiral strain with 1 particular antiserum are used to identify the presumptive serogroup of the strain (Dikken and Kmety, 1978).

Typing with monoclonal antibodies (mAbs) uses panels of mAbs that agglutinate serovars in a characteristic way. The mAbs recognize a small number of epitopes on the lipopolysachharides (LPS) which are shared by different serovars. The six isolates were further typed by performing MAT with a panel of monoclonal antibodies (WHO/FAO Collaborating Centre for Reference and Research; KIT-Biomedical Research) capable of differentiating the Icterohaemorrhagiae serovars (F20C3-1, F20C4-3, F52C1-4, F52C2-4, F70C4-5, F70C13-5, F70C20-4, F70C24-20, F70C26-3, F82C1-3, F82C2-3, F82C7-3, F82C8-7, F89C3-3, F89C12-6), Canicola (F152C1, F152C2, F152C5, F152C7, F152C8, F152C10, F152C11, F152C13, F152C14, F152C17, F152C18), Bataviae (F129C2-3, F129C3-3, F129C4-1, F129C6-1, F129C7-1, F129C9-5, F129C15-3, F129C18-1, F129C19-3, F129C20-5, F129C24-3, F129C25-1, F129C26-1).

2.4.2 Cross- agglutination absorption test

Cross agglutination absorption test (CAAT) is the basic test for characterisation of leptospires at the serovar level. Firstly, the cross agglutination test was performed to select serovars to be included in the CAAT. Then, the putative serogroup or groups to which the unknown strains belonged were determined by performing MAT with the

unknown strain as antigen and a panel of rabbit sera that is representative of all pathogenic serogroups. Strains might be allocated to more than one serogroup. Secondly, candidate serovars / strains with the positive serogroup(s) to be included in the CAAT are selected. This was carried out by performing MAT of the unknown strain with various rabbit antisera against serovars/strains within the positive serogroup(s). Antiserum against the unknown strain (i.e. 782 strain) was produced in a rabbit and tested in the MAT with all reference serovars/strains which belong to the positive groups. Cross-agglutination titers were expressed as percentages of the reciprocal titer for the homologous strain of positive group antisera with the unknown strain and vice versa. Two strains are considered to belong to different serovars if after cross-absorption with adequate amounts of heterologous antigen more than 10% of the homologous titer regularly remains in at least one of the two antisera in repeated tests (TSCL, 1987). The amount of antigen to be used for absorption is a very important factor (Wolff and Broom, 1954; Kmety et al., 1970). The antiserum was standardized to a MAT titre of 1:5,120. The absorbed antiserum whose titre with the absorbing strain approaches zero (and should not exceed 1% of the pre-absorption titre) is used for the agglutination with the homologous strain. Tests are repeated several times. The typing of leptospires at present takes a 10% limit criterion as decisive. This rule leaves a 0-10% margin of difference for strains belonging to the same serovar (Faine, 1982).

2.4.3 In silico DNA-DNA hybridization

Genome-to-Genome Distance Calculator (GGDC) is a state-of-the-art *in silico* method for genome-to-genome comparison, thus reliably mimicking conventional DDH. The DNA G+C content was calculated directly from the genome sequences. The resulting G+C content differences are compared with DNA–DNA hybridization (DDH) similarities calculated *in silico* using the GGDC web server, with 70% similarity as the gold standard threshold for species boundaries. The results indicate that the G+C content, if computed from genome sequences, varies no more than 1% within species. Taxonomic affiliation of all five *L. interrogans* strains was confirmed by pairwise comparison to *L. interrogans* ATCC43642^{T} (downloaded from https://www.ncbi.nlm.nih.gov/assembly/GCF_900156205.1) applying the genome-to-genome distances calculator (Meier-Kolthoff et al, 2012).

2.4.4 16s rRNA typing

16S rRNA sequences of whole genome sequences of six strains of *L. interrogans* were extracted using Artemis (Rutherford et al., 2000) in reference to primers LA (5' GGCGGCGCGTCTTAAACATG-3') and LB (5'TTCCCCCCATiGAGCAAGATT- 3'), which were described previously (Mérien et al.,1992). The sequences were then selected and aligned against 16S rRNA sequences of *L. biflexia* serovar Patoc and other ten *L. interrogans* strains available in the GenBank using CLUSTAL X (Gibson et al., 1994) in MEGA7 (Kumar, 1993).

Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach and then selecting the topology with superior log-likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site (next to the branches). The analysis involved 17 nucleotide sequences. All positions containing gaps and missing data were eliminated. There was a total of 327 positions in the final dataset. Evolutionary analyses were conducted in MEGA 7 (Kumar et al., 2016).

2.4.5 Multilocus sequence typing

The nucleotide sequences of seven housekeeping genes, i.e. *glmU, pntA, sucA, tpiA, pfkB, mreA* and *caiB,* were extracted from six *L. interrogans* whole genome sequences using Artemis (Rutherford et al., 2000) for multilocus sequence typing (MLST) (Scheme 1). The sequence types (STs) were assigned using the MLST website (http://leptospira.mlst.net/) (Boonsilp, 2013).

2.4.6 Core genome multilocus sequence typing

An ad-hoc cgMLST scheme was constructed in order to obtain the genome-wide gene-by-gene whole genome typing of *L. interrogans*. Target loci were determined by genome-wide gene-by-gene comparison using the cgMLST Target Definer (version 1.4) function of the SeqSphere+, 3.4.0 (Ridom GmbH, Münster, Germany), with default parameters (Jünemann et al., 2013). These parameters comprised the following filters to exclude certain genes of the *L. interrogans* serovar Lai strain 56601 reference genome (RefSeq NC_004342.2) from the cgMLST scheme: a minimum length filter that discards all genes shorter than 50 bp, a start codon filter

that discards all genes that contain no start codon at the beginning of the gene, a stop codon filter that discards all genes that contain no stop codon or more than one stop codon or if the stop codon is not at the end of the gene, a homologous gene filter that discards all genes with fragments that occur in multiple copies within a genome (with identity of 90% and more than 100-bp overlap), and a gene overlap filter that discards the shorter gene from the cgMLST scheme if the two genes affected overlap by >4 bp.

The remaining genes were then used in a pairwise comparison using BLAST (Altschul et al., 1997), version 2.2.12 (parameters used were the following: word size, 11; mismatch penalty, -1; match reward, 1; gap open costs, 5; gap extension costs, 2), with seven query genomes of *L. interrogans* available in NCBI RefSeq strains (i.e. *L. interrogans* serovar Copenhageni str. Fiocruz L1-130, *L. interrogans* serovar Hardjo str. Norma, *L. interrogans* serovar Manilae strain UP-MMC-NIID LP, *L. interrogans* serovar Manilae strain UP-MMC-NIID HP, *L. interrogans* serovar Bratislava strain PigK151, *L. interrogans* serovar Linhai str. 56609*, L. interrogans* serovar Lai str. IPAV and 15 draft genomes in Table A13) sequenced using MiSeq Personal Sequencer (Illumina Inc., San Diego, CA, USA). Using all genes of the reference genome that were common in all query genomes, with a sequence identity of ≥90% and 100% overlap and with the genome filters start codon filter, stop codon filter, and stop codon percentage filter turned on, the final cgMLST scheme was formed. Therefore, all genes having no start or stop codon in one of the query genomes, as well as genes that had internal stop codons in more than 20% of the query genomes, were discarded. The new ad-hoc scheme was developed using cgMLST target definer and evaluated on the five *L. interrogans* finished genomes and one draft genome of *L. interrogans.*

Thus, SeqSphere+ extracted the defined MLST+ core genome genes from each assembly with default parameters, mainly consisting of the following settings: (i) processing options: ‘Ignore contigs shorter than 200 bases'; (ii) scanning options: ‘Matching scanning thresholds for creating targets from assembled genomes' with ‘required identity to reference sequence of 90%' and ‘required aligned to reference sequence with 100%'; (iii) BLAST options: ‘Word size: 11', ‘Mismatch penalty: -1', ‘Match reward: 1', ‘Gap open costs: 5', and ‘Gap extension costs: 2'. In addition, the

MLST+ scheme genes were assessed for quality, i.e. the absence of premature stop codons, ambiguous nucleotides, and support of variants to the reference sequence by 75% or more read nucleotides (Jünemann et al., 2013).

2.4.7 Single nucleotide polymorphism (SNP) genotyping

Genome-wide single nucleotide polymorphism (SNP) patterns were analysed by using ParSNP v1.2 (Treangen et al, 2014). ParSNP is a core genome aligner which focuses on identifying the set of orthologous sequence conserved in all aligned genomes. FASTA files of genomes of interest were used as input and whole-genome alignment and mapping were performed against the first chromosome of the *L. interrogans* serovar Lai strain 56601 (RefSeq NC_004342.2). These commands were used for the analysis:

"cd Desktop/Parsnp-OSX64-v1.2/

./parsnp -p 4 -d Genome/ -g Lai.gbk"

-d = <path>: (d)irectory containing genomes/contigs/scaffolds

-r = <path>: (r)eference genome (set to to pick random one from genome dir)

-g = <string>: Gen(b)ank file(s) (gbk), comma separated list (default = None)

>parsnp -r <reference_genome> -d <genome_dir> -p <threads>

Parsnp adopts the compressed suffix graph to index the reference genome for identifying the maximum unique matches (MUMs) rapidly, which were used to recruit similar genomes and anchor the multiple alignments. By default, Parsnp calculates the maximum unique matches index (MUMi) distance between the reference and each of the genomes in the genome directory. All genomes with a MUMi distance ≤ 0.01 in the genome directory were included in the analysis. Gaps between collinear MUMs were aligned using MUSCLE (Edgar, 2004). The goal of Parsnp is to capture all informative signals found in the core genome of the specified clade of interest. Any SNPs in regions not shared by all genomes will not be reported. Additionally, any SNP found in a likely poorly aligned region would also be discarded. Finally, ParSNP does not perform LCB extension and therefore may miss SNPs appearing at the end of locally conserved blocks or clusters. Output was produced as multi-alignments (XMFA), variants (VCF), core genome phylogeny (Newick) and Gingr

input format (GGR). The Newick files were later exported into MEGA (Kumar et al., 2016) 7 for better resolution.

2.5 Genome Descriptions

2.5.1 Genome overview

General parameters of five complete genomes and one draft genome, i.e. size of chromosome 1 and 2, number of contigs in each genome, number of ribosomal RNA, tmRNA, tRNA, repeat region, gene, coding sequences (CDS), signal peptide and miscellaneous RNA, were documented.

2.5.2 Protein orthology

Clusters of orthologous proteins were generated using ProteinOrtho V5.16b (Lechner et. al, 2011). Analysis of the core genome was made based on the numbers of common proteins present in all six strains. Paralogous proteins were gathered in the same clusters and unique genes were calculated for each strain applying the `-singles` option. Orthologous proteins were visualized in a Venn diagram (http://www.interactivenn.net).

2.5.3 Genome synteny

An analysis for genome synteny applying large collinear blocks (LCB) was performed using Mauve 2.4.0 (Darling et al., 2004).

2.5.4 Plasmids

Additional contigs that do not belong to chromosome I and II were designated as plasmids. The faa files of the contigs were queried in Blastp for identification of the plasmids. Functional annotation of the orthologous group of each plasmid was searched using Eggnog 4.5.1 (Huerta-Cepas et al., 2016).

2.5.5 Prophages

Genomes were queried for the presence of phages using PHASTER (PHAge Search Tool Enhanced Release) at http:// phaster.ca (Arndt et al.,2016).

2.5.6 Clustered regularly interspaced short pallindromic repeats

Genome sequences were queried for Clustered Regularly Interspaced Short Pallindromic Repeats **(**CRISPR) elements using CRISPRCasFinder (https://crisprcas.i2bc.paris-saclay.fr/) (Couvin et al., 2018). Spacers of CRISPR arrays were checked for the presence of mobile genetic elements by the blastn search engine (https://blast.ncbi.nlm.nih.gov) (Altschul et al.,1997).

2.5.7 Insertion sequences

Genome sequences were compared against insertion sequences database using ISFinder (https://isfinder.biotoul.fr/https://isfinder.biotoul.fr/) (Siguier et al., 2006). Only the hit of equal to e-value: 0.0 was included in this study. Transposases of the *rfb* locus were queried by Blastp search (https://blast.ncbi.nlm.nih.gov) (Altschul et al., 1997).

2.6 Comparative Genomic Analysis

2.6.1 O-antigen

The LPS biosynthetic system (*rfb* locus) of all five strains was compared using EasyFig (Sullivan et. al, 2011) and color-coded similarly to Fouts et al, 2016.

Amino acids of each genome corresponding to lipid A and sialic acid biosynthesis as studied in Fouts et al, 2016 were queried against that of *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 using Blastp search (https://blast.ncbi.nlm.nih.gov) (Altschul et al.,1997). The locus tags and e-value were recorded.

2.6.2 Pathogenesis mechanisms

Protein sequences of each genome that associated with main pathogenesis functions such as motility, chemotaxis, adhesion to extracellular matrix, immunodominance, as studied in Fouts et al, 2016 were queried against that of *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 using Blastp search (https://blast.ncbi.nlm.nih.gov) (Altschul et al,1997). The locus tags and e-value were recorded.

2.6.3 Bacterial Virulence factors

Proteins of each genome were queried against 2602 sequences of proteins associated with experimentally verified virulence factors in Virulence Factor Database (http://www.mgc.ac.cn) (Chen et. al, 2016). Hits equal to or less than e-value = e^{-52} which represent the highest score of alignment i.e. ≥ 200 were considered to be significant.

2.6.4 Alginate biosynthesis

Proteins associated with structural biosynthesis, regulation and genotypic switching for *Pseudomonas aeruginosa* and were compared against six genomes using Blastp search (https://blast.ncbi.nlm.nih.gov) (Altschul et al.,1997).

2.7 Comparative transcriptomic analysis

Transcripts per million (TPM) was calculated from reads per kilobase of transcript, per million mapped reads (RPKM-1) using this formula: $[TPM(i) = (RPKM\text{-}1(i)/(\text{sum RPKM-1 of all transcripts}) * 10^6]$ (Li et al., 2009).

TPM values from three biological replicates of logarithmic phase of cultures from strain Langkawi and strain 1530^d were summarized by their mean values. T-test and $\log_2$ fold changes (logFC) were calculated. Differential gene expression was stated for those logFC values greater 1 or smaller -1 and p-values < 0.05.

3.0 METHODS: BIOMARKER IDENTIFICATION OF *LEPTOSPIRA SPP.*

3.1 Patient Serum Sample

3.1.1 Ethics statement

The study protocol NMRR-14-1687-23346(IIR) involved the use of patients' sera obtained for clinically indicated diagnostics to be used for research purposes and was approved by the Medical Research & Ethics Committee, Ministry of Health Malaysia, Malaysia. All human biosamples were anonymized.

3.1.2 Patients and serum samples

Sera from 34 patients who presented in the year 2015 with symptoms suggestive of acute leptospirosis were obtained from the Public Health Laboratory of Kota Bharu and Kota Kinabalu, Malaysia. The sera were obtained from hospitalized patients after an average of three to seven days of illness and had MAT titers of 1:800, which was consistent with the acute phase of leptospirosis. Infection with other pathogens, including Dengue virus, was excluded by the respective diagnostics as suggested by clinical history and other laboratory findings. Dengue fever is the most common infectious disease in this region and it may cause false positive results in *Leptospira spp.* ELISA tests due to cross-reactivity. All samples were therefore tested for the Dengue virus NS1 protein and anti-Dengue IgM and IgG reactivity and were thus ensured to be negative.

The patients presented with fever and one or more of these signs and symptoms: headache, myalgia, arthralgia, conjunctival injection, anuria or oliguria and/or proteinuria, jaundice, pulmonary and/or intestinal hemorrhage, cardiac arrhythmia or failure, skin rash, and gastrointestinal symptoms such as nausea, vomiting, abdominal pain, and diarrhea.

Sera from 18 patients were pooled into two groups based on their reactivity to Malaysian strains (n = 8) or WHO reference strains (n = 10). Two sets of Leptospira-negative control sera were used. The control pool used in Figure B3 was obtained from 7 healthy adult volunteers from Germany with no travel history to Leptospira-endemic countries. The other 16 patient sera were used for the titration ELISAs in the ORFeome procedure. The individual control sera for ELISA validation shown in Figure B4(A) were obtained from 16 healthy adults of Caucasian origin who

participated in an unrelated epidemiological study (Akmatov et al., 2011). Absence of leptospiral seroreactivity in the control sera were confirmed by in-house ELISA with leptospiral culture antigens.

3.2 Construction of *Leptospira spp.* genomic DNA library

3.2.1 Genomic DNA Library

Two libraries of *Leptospira spp*. were constructed using genomic DNA. The first library consisted of five strains of *L*. *interrogans* isolated from leptospirosis patients in Malaysia between 2014 and 2015. The second library consisted of seven WHO reference strains which were obtained from the Leptospirosis Reference Centre (also known as OIE Reference Laboratory for Leptospirosis, Amsterdam Medical Centre, Amsterdam). The strains are listed in Table B1. Strains from both groups were cultured in Ellinghausen-McCullough-Johnson-Harris (EMJH) medium at 30°C for 7–10 days at 250 rpm. Genomic DNA was isolated from pellets of 5 mL culture centrifuged at 8000 x g for 30 minutes (min), using the QiaAmp DNA Mini Kit according to the manufacturer's instructions (Qiagen, Hilden, Germany). The extracted DNA for each library was mixed and amplified with the illustra™ Ready-To-Go GenomiPhi V3 DNA amplification kit (GE Healthcare) according to the manufacturer's instructions. Twenty µg of DNA from each mixed and amplified genomic library were fragmented by sonication upon extraction. Subsequently, the DNA was concentrated using Amicon Ultra 0.5 mL centrifugal filters with a cut-off of 30 kDa.

Table B1. List of strains used to construct the two *Leptospira spp.* genomic libraries.

Genomic Library	Species	Serogroup	Serovar	Strain
I	*L. interrogans*	Canicola	Bindjei	782
	L. interrogans	Ictrohaemorrhagiae	Copenhageni	898
	L. interrogans	Bataviae	Paidjan	1489
	L. interrogans	Bataviae	Losbanos	1548
	L. interrogans	Icterohaemorrhagiae	Smithi	1530
II	*L. interrogans*	Australis	Australis	Ballico
	L. interrogans	Bataviae	Bataviae	Swart
	L. weillei	Celledoni	Celledoni	Celledoni
	L. kirschneri	Grippotyphosa	Grippotyphosa	Duyster
	L. interrogans	Sejroe	Hardjoprajitno	Hardjoprajitno
	L. borgpetersenii	Javanica	Javanica	Veldrat Batavia 46
	L. biflexia	Semaranga	Patoc	Patoc 1

DNA fragments with sizes from 100 to 800 bp were extracted from an agarose gel and the DNA ends were repaired with the Fast DNA End Repair Kit (Thermo Scientific) according to the manufacturer's instructions. 1.4 µg of fragmented DNA were then ligated into 1.4 µg of the *Pme*I-digested pHORF3 vector (Kügler et al., 2008) and subsequently transformed into *E. coli* TOP10F' (Invitrogen) by electroporation. Colony PCR was performed in some of the resulting clones to determine the insert rate of ligation.

3.2.2 Packaging of a phage display library with hyperphage

The library was packaged using Hyperphage (Rondot et al., 2001; Soltes et al., 2007) as described before (Hust et all., 2006; Kügler et al., 2008). By packaging the genomic DNA library with Hyperphage, ORFs are enriched and the resulting oligopeptides are presented on the phage particles for panning. The *E. coli* XL1-Blue MRF' containing the library was inoculated into 400 mL 2x YT-GA medium (2x yeast-tryptone broth supplemented with 0.1 M glucose and 100 µg/mL ampicillin) to an OD_{600} <0.1 and grown at 37°C, 250 RPM until OD_{600} ≈0.5. At this point, the culture was infected with Hyperphage (MOI 1:20) for 30 min at 37°C without shaking, and then 30 min under 250 RPM.

The culture was then centrifuged, suspended in 400 mL 2x YT-AK medium (2x YT containing 100 µg/mL ampicillin and 50 µg/mL kanamycin), and phage particles were produced at 30°C and 250 rpm overnight. Cells were then centrifuged for 20 min at 10,000 x g, and phage particles in the supernatant were precipitated with 1/5 volume of polyethylene glycol (PEG)/NaCl solution (20% w/v PEG 6000), 2.5 M NaCl) for 3 hours (h) on ice with gentle shaking. Phage particles were then pelleted for 1 h at 10,000 x g and suspended in 10 mL phage dilution buffer (10 mM TrisHCl pH 7.5, 20 mM NaCl, 2 mM EDTA). Remaining bacteria were pelleted by an additional centrifugation step of 10 min at 20,000 x g, and the solution was then filtered through a 0.45 µm filter to remove residual bacteria. The filtrate was again precipitated with 1/5 PEG/NaCl for 1 h and then centrifuged for 30 min at 20,000 x g. Pellets were suspended in 1 mL phage dilution buffer and residual bacteria removed by centrifugation for 1 min at 16,000 x g. The final supernatant containing the oligopeptide presenting phages was stored at 4°C. Phage titers were determined as

described previously (Hust et al., 2007).

3.2.3 Library validation by colony PCR and sequencing

To check library quality, a number of random *E. coli* colonies were analyzed by colony PCR and sequenced after packaging with Hyperphage. Therefore, the primers MHLacZPro_f (5'-GGCTCGTATGTTGTGTGG-3') and MHgIII_r (5'-GGAAAGACGACAAAACTTTAG-3') were used with the following PCR protocol: 98°C 30 s, 98°C 10 s, 56°C 20 s (35 cycles), 72°C 60 s and a final extension of 72°C for 2 min. The DNA was separated and analyzed by gel electrophoresis (Qiaxcel Advanced). Additionally, plasmid DNA was sequenced with the primers used for colony PCR to verify the correct inserts and the ORF enrichment after packaging.

3.3 Selection of Immunogenic Oligopeptide Phage by Panning

3.3.1 Panning

Two wells of a Maxisorp™ 96 well microplate were coated with 150 µL of goat IgG directed against human IgG, IgA and IgM Fc (Dianova 109-005-064, 2 mg/mL, Lot No. 113036) diluted in 1:500 phosphate buffered saline (PBS) and another six wells were coated with 5 x 10^{10} CFU Hyperphage in PBS and incubated at 4°C overnight. Subsequently, the coating solutions were removed and the wells were blocked for 30 min with PBS with 0.1% Tween, supplemented with 2% (w/v) milk powder (2% MPBST).

A pool of twelve patients' sera was diluted 1:100 in 2% MPBST and pre-incubated twice in the Hyperphage coated wells for 1 h to eliminate IgG binding to helper phage. After pre-incubation, the serum pool was incubated for 1.5 h in the wells coated with goat anti-human IgG, A, M antibody. The *Leptospira spp*. phage library (corresponding to 1.1 x 10^{10} CFU) was mixed 1:3 with 2% MPBS-T and incubated in the wells with the pooled leptospirosis patients' sera for 1.5 h. Unbound phage and phage with low affinity were removed by stringent washing steps. Three panning rounds were performed, and the wells were washed twice after each step, with one additional wash for the second panning round. After washing, bound phage particles were eluted with 200 µL of 10 µg/mL trypsin in PBS for 30 min at 37°C. Eluted

phages of both wells were combined and 10 µL were used for titration. The remaining 390 µL were used to infect 20 mL of an *E. coli* TOP10F' culture grown to an OD_{600} of 0.5. The cells were incubated for 30 min at 37°C and harvested by centrifugation for 10 min at 3250 x g. The pellet was suspended in 250 µL 2xYT-GA. The bacterial suspension was plated onto 15 cm 2xYT-GA agar plates and incubated overnight at 37°C. Colonies were swept off with 5 mL 2xYT-GA medium, then 50 mL of 2xYT-GA medium was inoculated with the bacterial suspension to an OD_{600} of <0.1 and grown to an OD_{600} of 0.5 at 37°C and 250 rpm. For infection, 5 mL of the bacterial culture (approx. 2.5 x 10^9 cells) was mixed with Hyperphage infected at an MOI of 1:20 resulting in 5 x 10^{10} CFU Hyperphage. The suspension was incubated at 37°C for 30 min without shaking and another 30 min at 37°C and 250 rpm. To remove glucose, which inhibits phage expression, the infected cells were harvested by centrifugation for 10 min at 3220 x g. The remaining pellet was resuspended in 30 mL 2x YT-AK and incubated at 30°C and 250 rpm overnight for phage production. The bacterial cells were then pelleted by centrifugation for 20 min at 3220 x g and the remaining supernatant was used to precipitate phage particles with PEG/NaCl (20% (w/v) PEG 6000, 2.5 M NaCl). Thirty mL of supernatant were separated and incubated for 1 h with 6 mL PEG/NaCl on ice with slight shaking on a rocker, followed by centrifugation at 6000 x g for 1 h at 4°C. The phage pellet was resuspended in 500 µL phage dilution buffer (10 mM TrisHCl pH 7.5, 20 mM NaCl, 2 mM EDTA), centrifuged in a microcentrifuge at 16,100 x g for 1 min and the supernatant was used for further panning rounds. For the 2nd and 3rd panning rounds, 150 µL of the amplified phage was used. Eluted phage particles from the 3rd panning round were used for titration without further amplification. Single colonies were then used for single oligopeptide phage production.

3.3.2 Production of single oligopeptide phage clones for screening

Single oligopeptide phage clones were produced by inoculating 175 µL 2x YT-GA medium with single colonies from the titration plate in a polypropylene 96-well U-bottom plate (Greiner bio-one). The cultures were incubated at 37°C and 500 rpm shaking overnight.

From this plate, 10 µL were used to inoculate another 165 µL 2xYT-GA medium per well, which was incubated at 37°C and 800 rpm for 2 h. Subsequently, the bacteria

were infected with 5 x 10^9 cfu Hyperphage and incubated for 30 min at 37°C without shaking and 30 min at 37°C and 800 rpm. The bacteria were pelleted by centrifugation at 3220 x g for 10 min and the pellets were resuspended in 175 µL/well 2x YT-AK and incubated overnight at 30°C and 800 rpm. The produced phage in the supernatant were transferred to another plate and precipitated with 1/5 volume of PEG/NaCl solution for 1 h at 4°C. Next, precipitated phage particles were pelleted by centrifugation at 3220 x g for 1 h and the pellets dissolved in 150 µL PBS. Remaining bacterial cells were separated by another centrifugation step and the phage-containing supernatants stored in a new plate at 4°C and used for screening ELISA.

3.4 ELISA of Oligopeptide Phage Clones

3.4.1 Screening ELISA

Two types of ELISA were performed for selection of oligopeptide phage clones. First, a screening ELISA was performed on each genomic library. Oligopeptide phage particles were captured by monoclonal mouse anti-M13 (B62-FE2, Progen) antibody for screening. For this, 100 µL of a 250 ng/mL solution of antibody in PBS were coated overnight at 4°C and subsequently blocked with 2% MPBST. The wells were washed after each incubation step three times with 300 µL PBST. One hundred µL of the monoclonal phage clones were added to each well and incubated for 2 h at 4°C. 100 µL of pooled patients' sera reactive to Malaysian strains were added to library 1 while pooled sera reactive to WHO strains were added to library 2. All sera were diluted in 2% MPBST supplemented with 10% *E. coli* TOP10F' lysate and 1 x 10^{10} CFU/mL Hyperphage. The dilutions were incubated at RT for 2 h prior to use in the ELISA. Then, the dilutions were added onto the captured phage particles for 1.5 h and detected via a goat anti-human IgG, A, M antibody conjugated to horseradish peroxidase (HRP) (1:20,000) for 1.5 h. Visualization was achieved by adding 100 µL TMB (3,3',5,5'-tetramethylbenzidine) solution and the reaction was stopped with 100 µL 1 N sulfuric acid. A SUNRISE microtiter plate reader (Tecan, Crailsheim, Germany) was used to measure absorbance at 450 nm and subtract scattered light at 620 nm.

3.4.2 Titration ELISA

A second titration ELISA was performed on the selected unique oligopeptide clones from the screening ELISA. The method was the same except this time clones were tested for reactivity against each of three aforementioned serum pools. The sera were serially diluted 2-fold from 1:100 to 1:102,400 in 2% MPBST supplemented with 10% *E*. *coli* TOP10F' lysate and 1 x 10^{10} CFU/mL Hyperphage.

3.5 Validation of Immunogenic Peptides

3.5.1 Production of control proteins and synthetic peptide

The *lipL32* and *loa22* genes were amplified from *L*. *interrogans* serovar Copenhageni genomic DNA. Phusion DNA Polymerase (Thermo Scientific F-530L) was used to amplify full-length genes according to the following protocol: 98°C 30 seconds (s), 98°C 10 s, annealing temperature primer dependent 20 s, 72°C 20 s, 30 cycles, 72°C 10 min. The amplified genes were digested with Nde1 and Not1 and the resulting fragments resolved by 1% agarose electrophoresis, purified from the gel with NucleoSpin Gel and PCR Clean-Up kit (Macherey-Nagel 740609.250), ligated into the Nde1/Not1 digested vector pET21a(+) and the ligation product subsequently transformed into *E*. *coli* BLR (-DE3). Finally, positive clones were identified by colony PCR and confirmed by sequencing.

For protein expression, 200 mL 2xYT-GA medium were inoculated with 10 mL overnight culture and incubated at 37°C and 120 rpm in a baffled flask to an OD600 of 0.6. Expression was induced with a final concentration of 1 mM IPTG for 6 h at 30°C, followed by centrifugation at 3,000 x g for 20 min for cell harvesting. Cells were then suspended in His-tag binding buffer pH 8 with urea (50 mM Na2HPO4, 100 mM NaCl, 10 mM imidazol, 8 M urea) and incubated for 1 h under over-head rotation, followed by sonication (6 cycles of 10 s 50% power, 10s incubation on ice, Sonotrode MS72, Bandelin). Subsequently 0.5 mL Ni-NTA agarose slurry (Qiagen 30210) was added to the disrupted cell solution and incubated for 1 h under over-head rotation. Then, the solution was loaded onto a polypropylene column. The agarose settled by gravity flow and was washed with 10 mM, 30 mM and 50 mM

imidazole (50 mM Na2HPO4, 300 mM NaCl, 8 M urea, pH 8). Elution was achieved with 3 x 1.25 mL PBS pH 7.4 supplemented with 100 mM EDTA and 8 M urea.

The proteins were prepared for analysis by 12% SDS-PAGE by heating 0.5 µg of protein sample mixed with 5x Lane Marker Reducing Sample Buffer (Thermo Scientific 39000) at 95°C for 5 min. PageRuler Plus Prestained Ladder (Thermo Scientific 26619) and Spectra Multicolor Low Range Protein Ladder (Thermo Scientific 26628) were used as size marker. The samples were stacked for 10 min at 60 V, followed by separation for 60 min at 110 V. The gels were then stained with Coomassie Brilliant Blue R250 solution dye. The protein bands migrated at ~32 kDa and ~22 kDa.

Leptospira spp. whole-cell antigen for the whole-cell in house ELISA was prepared using the supernatant of live *Leptospira spp*. cultures. The antigen was prepared and coated to microtiter plates essentially as described by (Goris et al., 2011).

3.5.2 ELISA for validation of immunogenicity of synthetic peptides

To confirm the interaction of each isolated peptide with the sera, 200 ng of synthetic peptide (Peps4LS GmbH, Heidelberg, Germany) was diluted in 100 µL of PBS and coated to a high binding 96-well microtiter plate (Greiner-bio one) and incubated at 4°C overnight. Blocking was performed with 2% MPBST for 30 min. Then, 16 sera from patients with acute leptospirosis (MAT titer, 1:800; reactive towards endemic serovars in Malaysia i.e. Australis, Autumnalis, Bataviae, Canicola, Celledoni, Copenhageni, Djasiman, Gryppotyphosa, Hardjobovis, Hardjoprajitno, Icterohaemorrhagiae, Javanica, Lai, Patoc, Terengganu, Sarawak) and 16 sera from healthy donors were serially diluted 2-fold from 1:100 to 1:102,400 in 2% MPBST and added into the wells for 1.5 h. After incubation, the wells were washed three times with 300 µL PBST. Bound IgM was detected with mouse IgG anti-human IgM (CH2)-HRPO, MinX none 100 µg; product no. AFC-5349-2, Dianova) diluted 1:20,000 in 2% MPBST, by incubation for 1.5 h at RT. The reaction was developed with TMB solution, stopped with sulfuric acid, and the plates were read at 450 nm as described before.

Similar ELISA steps were repeated with control proteins (rLipL32 and rLoa22) and *Leptospira* antigens. For ELISA of control proteins, the sera were diluted and titrated from 1:100 to 1:102,400 in 2% MPBST supplemented with 10% *E*. *coli* TOP10F' lysate and 1 x 1010 CFU / mL Hyperphage and incubated for 2 h prior to use, as done in the ELISA of oligopeptide phage clones described above.

3.6 Statistical analyses

Antigen/antibody ELISA signals in control and disease groups were non-normally distributed, and statistical significance of between-group differences was therefore assessed with the Wilcoxon rank sum (Mann-Whitney U) test (Bauer, 1972). To evaluate discriminatory biomarker potential, a logistic regression model fitted using Bayesian generalized linear models (Gelman et al., 2008) was used to calculate the area under the receiver operating characteristic (ROC) curve (AUC). The AUC and corresponding confidence intervals (CI) values were estimated using the cross-validation procedure based on 1000 bootstrap samples as described before (DeLong et al., 1988; Zaki and Meira, 2014). In addition, multiple logistic regression was used to evaluate the classification performance of each combination of peptides / proteins.

3.7 List of Materials

Isolation of Genomic DNA	• DNA isolation kit
Amplification of Genomic DNA	• Illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit (GE Healthcare, Freiburg, Germany). • DNA-free water. • PCR reaction tubes • Thermocycler • Agarose (Peglab, Erlangen, Germany). • TAE-buffer 50x: 2 M Tris-HCl, I M acetic acid, 0.05 M EDTA pH 8. • Electrophoresis chamber.
Fragmentation of DNA	• Sonoplus HD220 and MS72 Sonotrode (Bandelin, Berlin, Germany). • Amicon Ultra Centrifugal Filters (30K) (Merck Milipore, Tullagree, Ireland). • Gel and PCR purification kit. • Agarose.

	• TAE-buffer 50x • Electrophoresis chamber
DNA End Repair	• Fast End Repair Kit (Thermo Fischer Scientific, Waltham, USA). • Gel and PCR purification kit.
Library Construction	• pHORF3 • PmeI-HF (NEB, Frankfurt am Main, Germany). • CIP (NEB, Frankfurt am Main, Germany). • CutSmart Buffer (NEB, Frankfurt am Main, Germany). • NucleoSpin Gel and PCR clean-up (Macherey-Nagel, Düren, Germany) • T4 DNA ligase (Promega, Mannheim, Germany) • Amicon Ultra Centrifuga Filters (30K) (Merck Millipore, Tullagree, Ireland). • Glycerol • 0.1cm electroporation cuvette. • Electroporator • SOC medium pH 7.0: 2% (w/v) tryptone, 0.5% (w/v)yeast extract, 0.05%(w/v) NaCl, 20mM Mg solution, 20mM glucose (sterilize magnesium and glucose separately, add solutions after autoclavation). • Polystyrene dish with lid (245mm x 245mm x 25mm). • 2xYT-GA agar: 2xYT, 100mM glucose, 100 µg/mL ampicillin 1.2% (w/v) agar-agar. • Electrocompetent *E.coli* TOP10F' (Invitrogen, Carlsbad, USA) (F'[lacIq Tn10(tetR)] mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZ ΔlacX74 deoR nupG recA1 araD139 Δ(ara—leu)7697 galU galK rpsL(StrR) endA1 λ-). • Liquid Nitrogen. • Single-use Drigalsky spatulas. • 2mL cryo vials • 10cm Petri dishes. • Agarose. • TAE-buffer 50x. • Electrophoresis chamber. • Optional: QIAxcel Advanced System (QIAGEN, Hilden, Germany).
Antigen Library Packaging	• 2xYT medium pH 7.0 • 2xYT: 2xTY + 20 µg/mL tetracycline • 2xYT-GA:2xYT + 100mM glucose + 100µg/mL ampicillin.

- 2xTU-GA agar plates
- 2xYT-AK: 2xYT + 100 µg/mL ampicillin + 50 µg/mL kanamycin.
- 10 cm Petri dishes.
- Hyperphage (Progen, Heidelberg, Germany).
- 1mL cuvettes and spectrophotometer 600nm wavelength
- 100 and 1000 mL glass shake flasks.
- 50 mL tubes.
- 0.45 µM syringe filters
- Syringe.
- Incubator for shake flasks.
- Eppendorf centrifuge (Eppendorf, hamburg, Germany).
- Sorval Centrifuge RC5B Plus, rotor F9S and SS34 (Thermo Fischer Scientific, Waltham, USA) and respective tubes.
- Polyethylenglycol-Sodium Chloride (PEG-NaCl) solution: 20% (w/v) PEF 60000, 2.5M NaCl.
- Phosphate-buffered saline (PBS) pH 7.4:8.0 g NaCl, 0.2 g KCl, 1.44g $Na_2HPO_4.2H_2O$,0.24g KH_2PO_4 in 1L.

Colony PCR

- Oligonucleotide primer:

MHLacZ-Pro_f GGCTCGTATGTTGTGTGG
MHgIII_r CTAAAGTTTTGTCGTCTTTCC

- GoTaq DNA Polymerase and buffer (Promega, Frankfurt am Main, Germany).
- dNTP Mix
- DNA-free H_2O
- Thermocycler

Antigen Panning and Screening

- 96-well ELISA Costar plate (Corning, Corning, USA).
- Phosphate-buffered saline (PBS) pH7.4
- PBS-T: PBS + Tween 20 0.05% (v/v).
- 2% MPBS-T: skimmed milk powder 2% (w/v) diluted in PBS-T.
- Tecan plate washing machine.
- Electrocompetent *E.coli* TOP10F' (Invitrogen, Carlsbad, USA) (F'[lacIq Tn10(tetR)] mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZ ΔlacX74 deoR nupG recA1 araD139 Δ(ara—leu)7697 galU galK rpsL(StrR) endA1 λ-).
- 1 mL cuvettes and spectrophotometer 600 nm wavelength.

- Trypsin (1 mg/mL stock).
- Eppendorf centrifuge
- 2xYT medium pH 7.0
- 2xYT-T
- 2xYT-GA
- 2xYT-GA agar plates
- 2xYT-AK
- 96-well U-shaped polypropylene plate.
- Hyperphage (Progen, Heidelberg, Germany)
- PEG-NaCl
- Anti-Ig Fc-specific anti HRP-conjugated
- Anti-M13 phage (pVIII) HRP-conjugated (GE Healthcare, Freiburg, Germany).
- TMB solution
- 1 NH_2SO_4

4.0 RESULTS: WHOLE GENOME ANALYSES OF *LEPTOSPIRA SPP.*

4.1 Clinical Cases and Investigations

Six isolates of *L. interrogans* were isolated from six patients; five from patients who were hospitalized in Hospital Tuanku Ampuan Rahimah (HTAR), Klang, Malaysia between 2014 and 2015, while the sixth isolate was isolated in the Netherlands in 2004 from a Dutchman who contracted leptospirosis while he was in Langkawi island in Malaysia on holiday. All five patients were male and their age range from 15 to 27 years old, while the Dutch patient was 50 years old.

The Dutch patient had history of waddling in the mangrove forest in Langkawi five days prior to the initial symptoms. Three patients had a history of swimming in a river in Selangor state prior to infection. The patient of strain 782 swam in a river in Ulu Yam 1 week prior to infection, the patient of strain 1489 swam in a river in Ulu Langat 2 week prior to infection and patient of strain 1530d had a history of jungle tracking and swimming in a river in Kemensah, Selangor. The patient of strain 898 was a Bangladeshi national who worked in a grocery shop with known history of rodent infestations. The patient of strain 1548 denied any recreational or sport history, and did not have any history of contact to any rodents or animals. The locations of the recreational places are shown in Figure A1.

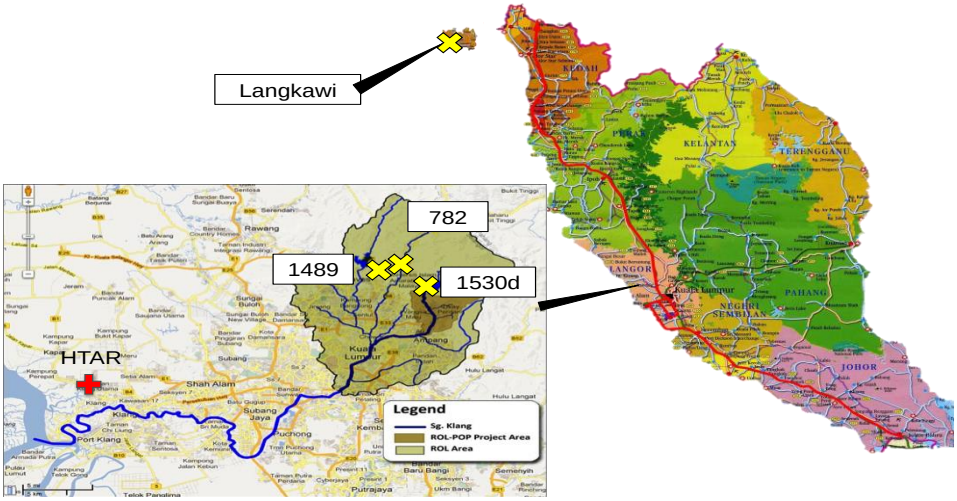


Figure A1. Location of infection origins of four leptospirosis cases in Peninsular Malaysia. Langkawi island is a resort island which retained natural attractions such as mangrove forests, where patient of strain Langkawi acquired the infection. Three cases went for swimming and recreation at three different rivers in Selangor.

Four out of six patients, i.e. patients of strains Langkawi, 782, 1489 and 1548, presented with acute symptoms of septic shock (i.e. high body temperature, chills, rigors, dyspnea, low blood pressure, raised heart rate and respiratory rate) and required resuscitation in the intensive care unit (ICU). The longest ICU stay was 4 days i.e. patient of strain 782, followed by 3 days by patient 1548. The patient of strain Langkawi was only admitted in the ICU for 2.5 days but was hospitalized for the longest duration among the six cases, i.e. for 11 days. The other five patients were hospitalized between 4 to 7 days before being discharged home.

All patients presented with fever, and most were associated with chills, rigors muscle pain (myalgia), arthralgia (joint pain) and gastrointestinal symptoms such as vomiting, epigastric pain and diarrhea. Patient of strains Langkawi and 782 had conjunctival injection and yellowish discoloration of sclera. Patient of strain 782 and strain 1489 had the lowest blood pressure suggestive of circulatory collapse on presentation, i.e. 88/55 mmHg and 88/49 mmHg, respectively, while the other patients presented with normal blood pressure.

Patient of strain Langkawi presented with high fever (39.5°C), tachycardia (pulse rate 124 /min) oxygen saturation was 96%, both scleras was injected without jaundice and blood pressure of 120/68 mmHg which deteriorated rapidly. The patient of strain Langkawi had hemoptysis, chest x-ray revealed signs of pulmonary haemorrhage (Figure A2), and the patient required blood transfusion due to severe hemoptysis (Wagenaar et al, 2005). These findings were consistent with severe pulmonary haemorrhage syndrome (SPHS) and associated with poor prognosis (Park et al., 1989, Trevejo et. al., 1998).

Figure A2. Chest X-ray of patient with strain Langkawi showed patchy opacification compatible with pulmonary hemorrhage (Wagenaar et al., 2005).

Additional symptoms were peripheral injection of the sclera (no signs of jaundice), macular exanthema, myalgia, nausea, vomiting, and diarrhea. All clinical findings were summarized in Table A1.

Table A1. Summary of patients' details: clinical histories, physical findings and complications in six leptospirosis patients.

	Langkawi *	**782**	**898**	**1489**	**1548**	**1530d**
Age	50	15	21	24	27	24
Gender	Male	Male	Male	Male	Male	Male
Admission (days)	11	6	6	4	7	6
ICU (days)	2.5	4	-	1	3	-
History of exposure	Wading mangrove forest	Swim in river	Rodent infestation at workplace	Swim in river	Deny any exposure	Jungle track & swim in river
Period of exposure	<1 week	1 week	nil	2 weeks	nil	3 weeks
Signs & Symptoms	Fever	High grade fever	Fever	High grade fever	Fever	Fever
	NR	Chills, Rigors	Chills, Rigors	Chills	Chills	Chills, Rigors
	Myalgia	Myalgia Arthralgia	NR	Myalgia Arthralgia	Myalgia	NR
	Nausea, Vomiting Diarrhea	Epigastric pain	Vomiting	Vomiting	Vomiting Epigastric pain Diarrhea	Vomiting
	Jaundice	Jaundice	Jaundice	NR	NR	NR
	Subconjunctival haemorrhage	Subconjunctival haemorrhage	Tea-coloured urine	NR	NR	NR
	Dyspnoea	Dypsnoea	NR	NR	Dyspnoea	NR
	Hemoptysis	NR	NR	NR	NR	NR
Complications	Septic Shock	Septic Shock	NR	Septic Shock	Septic Shock	NR
	Acute kidney injury	Acute kidney injury	Acute kidney injury	NR	NR	NR
	Pulmonary hemorrhage	NR	NR	NR	NR	NR
Isolation date/place	The Netherlands, 2004	Klang, Malaysia 2014	Klang, Malaysia 2014	Klang, Malaysia 2014	Klang, Malaysia 2015	Klang Malaysia 2014

ND, no data

NR, not relevant

The laboratory findings showed that patients of strain Langkawi, 782, 898 and 1530d presented with critical parameters suggestive of severe leptospirosis on arrival to the emergency department, which was white blood cell > 11 x 10^8 cells / L. Hemoglobin < 12 g/dL and platelet <100 x 10^9/L was reported on patient 782. Serum creatinine values >120 µmol/L were reported on patients strain Langkawi, strain 782 and 898; and blood urea >10 µmol /L was reported on patient strain 898. The patient of strain

782 had deranged coagulation profile illustrated by results of patient prothrombin time (PT) and activated partial thromboplastin time (APTT), i.e. 15.4 and 60.9, respectively, and was transfused with fresh frozen plasma which led to normalization of the parameters. C-reactive protein (CRP), an acute phase serum protein, was raised in four cases, i.e. 782, 898, 1489 and 1548. There was no data on this test for strain Langkawi and 1530d. Liver function of strain Langkawi and 782 was affected as total bilirubin level raised, alanine transferase (ALT) and aspartate transferase (AST) increased to 370 U/L and 1,110 respectively for strain Langkawi; and 62 U/L and 59 U/L respectively for strain 782. Serum bilirubin was raised in the patient of strain Langkawi, 782 and 898, consistent with the jaundice seen in these patients.

The patients of strain Langkawi and 782 had presentations similar that resembled that of Weil's disease with peripheral scleral injection sclera, jaundice and acute kidney injury (Weil,1886). Both patients also presented with septic shock that required resuscitation and mechanical ventilation in the ICU. Three cases, i.e. the patients of strain 782, strain 898 and strain 1548 showed signs of abrupt decline in renal functions evidenced by raised nitrogen waste products i.e. serum urea and creatinine and oliguria.

Only the patient of strain Langkawi was confirmed with a diagnosis of leptospirosis during management in the ward with a positive *Leptospira* IgM ELISA of 1: 160, which increased to 1: 5,120 on day 14 and was confirmed later by MAT. Four out of six patients were treated with a combination of intravenous (IV) doxycycline and ceftriaxone, which are the second line combination antibiotics of choice for leptospirosis treatment in Malaysia (National Antibiotic Guideline, 2014). In the Dutch hospital, the patient of strain Langkawi was treated with doxycycline as the empirical antibiotic (as to cover both leptospirosis and the differential diagnosis of rickettsiosis), and this was later changed to amoxycillin after laboratory results were out. The patient of strain 1548 was treated with amoxycillin-clavulanate and azithromycin as he had symptoms of a headache, vomiting and shortness of breath which suggestive of allergy to ceftriaxone, and also suspected of having atypical pneumonia as the anti-mycoplasma test was positive with a titer of 1:40.

All patients including patient of Langkawi strain were discharged well and all Malaysian patients were treated with oral antibiotics i.e. doxycycline and cefuroxime

to complete the treatment course. A follow-up chest x-ray of the Langkawi patient showed no pulmonary sequelae.

The patient of strain Langkawi had the highest results for specific parameters i.e. white blood cell count, serum creatinine, liver enzymes and total bilirubin and creatinine kinase, which indicated severe leptospirosis. The patient of strain 782 had the highest CRP level and other parameters indicating severe leptospirosis. The patients of strain 1548 and 1530d results indicated milder disease status (Table 2).

Table A2. Summary of clinical parameters, laboratory results and treatment in six leptospirosis patients.

Parameter	Langkawi	782	898	1489	1548	1530d
WBC count (>11 x 10^9 cells/L)*	16.2	14.9	11.7	11.0	6.5	15
Haemoglobin (<12g/dL)*	ND	11.6	14.6	13	11.2	14
Platelet count (<100 x10^9/L)*	115	31	127	181	105	184
PT (11 to 13.5 s)	13.7	15.4	ND	ND	ND	ND
APTT (30 to 40 s)	ND	60.9	ND	ND	ND	ND
INR (0.8 to 1.1)	ND	1.28	ND	ND	ND	ND
C-reactive protein (0.8 mg/L to 3.0 mg/L)	ND	2161	165	93	264	ND
Serum creatinine (>120µmol/L)*	315	124	171	ND	ND	ND
Blood urea (>10mmol/L)*	ND	9.1	10.1	ND	ND	ND
Serum sodium (130-150 mmol/l)	ND	136	138	ND	ND	ND
Serum potassium (3.5 to 5.5 mmol/L)	ND	3.6	3.9	ND	ND	ND
ALT (>45 U/L)*	370	62	22	ND	ND	ND
AST (>40 U/L)*	1110	59	80	ND	ND	ND
CK (>190 U/L)*	5,421	280	262	ND	ND	ND
Serum bilirubin (17µmol/L)*	ND	29	ND	ND	ND	ND
Total bilirubin (mg/dL)	258	ND	65.3	ND	ND	ND
SBP mmHg on arrival	120	88	102	88	130	101
DBP mmHg on arrival	68	55	62	49	85	60
Pulse rate (min)	124	140	100	79	90	96
Respiratory rate on arrival	ND	35/min	35/min	35/min	35/min	35/min
Glasgow coma scale	ND	15/15	15/15	15/15	15/15	15/15
Inotropes (maximum dose)	ND	*Noradr 8mg	NR	*NorAdr 4mg	NR	NR
Dopamine	yes	no	NR	no	NR	NR
Oxygen saturation (SpO2)	96%	98%	98%	99%	93%	NR
Temperature (on arrival)	39	39	38.5	38	39	NR
1st antibiotics	Amoxycillin	Doxycycline	Doxycycline	Doxycycline	Augmentin	Doxycycline
2nd antibiotics	Doxycycine	Rocephine	Rocephin	Rocephin	Azithromycin	Rocephin
Leptospira IgM	1 in 160	Negative	Negative	Negative	Negative	Negative
Dengue IgM	NR	Negative	Negative	Negative	Negative	Negative
Anti-mycoplasma IgM	NR	NR	NR	NR	1 in 40	NR

ND, no data
NR, not relevant
ALT, alanine transferase; AST, aspartate trasferase; CK, creatinine kinase; WBC, white blood cells; PT, prothrombin time; APTT, activated partial thromboplastin time; INR, international normalized ratio; *Parameters associated with severe leptospirosis (Marasinghe et al., 2017).

Based on this information, the severity of disease of these patients could be classified as follows: patient of strain Langkawi had the most severe leptospirosis, followed by patient of strain 782. On the other hand, patients of strains 898 and 1530d are two cases without admission to the ICU. Comparing the two, patient of strain 1530d had the mildest leptospirosis with milder clinical presentations and

without any complications. Patients of strains 1489, 1548 and 898 can be classified as mild to moderate leptospirosis cases. It was thus assumed that the degree of virulence of the corresponding isolates differed accordingly.

4.2 Virulence Test in Animal Model

Based on the survival test of guinea pigs, two guinea pigs infected with strain 782 showed the earliest (day 3) signs requiring euthanasia, i.e. dehydration, weight loss, decreased activity, isolation, and icterus and were euthanized on the following day. Another guinea pig infected with strain 782 only showed distress signs on day 6, and it was euthanized on the following day. On day 5, one of the guinea pigs in strain Langkawi group showed signs of distress, followed by two other guinea pigs on the following day. All guinea pigs in strain 1548 and 1530d groups consistently showed clinical signs starting on day 7 and were euthanized two days after that. Guinea pigs in the strain 1489 group showed an inconsistent survival pattern, as the first guinea pig died on day 6, the second only showed signs on day 13 and the third one earlier seven days. The last group to be euthanized was group 898, i.e. on day 14 after inoculation. Table A3 summarizes survival (in days) of the guinea pigs and Figure A3 shows weight change patterns.

Table A3. Days of survival in seven groups of guinea pigs i.e. one control and six groups infected with six strains of *L. interrogans.* Guinea pigs with signs and symptoms of weight loss, ruffled fur, and isolation (WHA) were euthanized. Sera and organs for histology samples were taken.

Group strain	Guinea Pig No	Physical Changes*	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
1. Control	Guinea Pig 1	-															E
	Guinea Pig 2	-															E
	Guinea Pig 3	-															E
2. 782	Guinea Pig 1	W, H, A			+	E											
	Guinea Pig 2	W, H, A						+	+	E							
	Guinea Pig 3	W, H, A			+	E											
3. 898	Guinea Pig 1	W, H, A													+	+	E
	Guinea Pig 2	W, H, A													+	+	E
	Guinea Pig 3	W, H, A													+	+	E
4. 1489	Guinea Pig 1	W, H, A						D									
	Guinea Pig 2	W, H, A													+	+	E
	Guinea Pig 3	W, H, A							+	+	E						
5. 1530	Guinea Pig 1	W, H, A							+	+	E						
	Guinea Pig 2	W, H, A							+	+	E						
	Guinea Pig 3	W, H, A							+	+	E						
6. 1548	Guinea Pig 1	W, H, A							+	+	E						
	Guinea Pig 2	W, H, A							+	+	E						
	Guinea Pig 3	W, H, A							+	+	E						
7. Langkawi	Guinea Pig 1	icterus						+	+	E							
	Guinea Pig 2	icterus					+	+	E								
	Guinea Pig 3	icterus						+	+	E							

+ dehydration, weight loss (W), ruffled hair coat (H), decrease activity and isolation (A); E, euthanasia; D, death

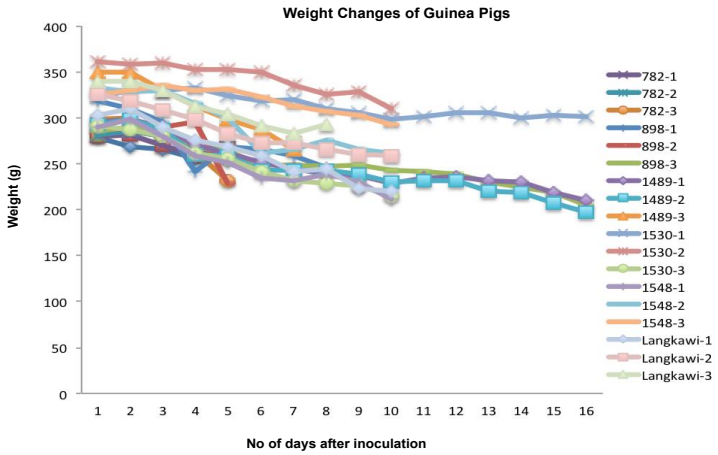


Figure A3. Weight change patterns in six groups of guinea pigs infected with the six strains of *L. interrogans*. Generally, all guinea pigs showed weight reduction between day 5 to day 15 after inoculation.

Biochemistry test of the guinea pigs sera were investigated and the results are displayed in Table A4. This result showed that guinea pigs 1 and 3 inoculated with strain Langkawi had raised total bilirubin (i.e. 4.02 mg/dL and 7.47 mg/dL, respectively). However, it was unclear whether conjugated or unconjugated bilirubin contributed to the raised total bilirubin. Furthermore, guinea pigs infected with strain Langkawi had elevated serum creatinine level (0.84 mg/dL and 0.40 mg/dL), which was significantly higher than in the other groups and suggestive of acute kidney injury. There was raised in serum AST in guinea pig 2 of strain 898 and guinea pig 1 of strain 1548 compared to another group of guinea pigs. However, these findings were not consistent with the other 2 guinea pigs in each group and, thus, are relatively inconclusive.

Table A4: Serum biochemistry test results of one control and six infected groups of guinea pigs. There were three guinea pigs in each group however for group 2 and 7, only two guinea pigs had enough serum sample for biochemistry testing. There are no reference values for these tests for guinea pigs, thus, these results should be interpreted relative to the control group.

Group No	Guinea pig No	ALT U/l	ALP U/l	Total bilirubin mg/dl	Direct bilirubin mg/dl	Indirect bilirubin mg/dl	Urea mg/dl	Creatinine mg/dl	Total Protein g/dl	Albumin g/dl	Amylase U/l
	Control 1	9	28	0,00	0,00	0,00	43	0,37	4,21	2,32	579
1	Control 2	8	17	0,00	0,00	0,00	45	0,31	4,14	2,15	855
	Control 3	11	27	0,01	0,00	0,01	45	0,39	4,57	2,45	943
2	782-1	9	25	0,01	0,00	0,01	19	0,39	4,67	2,37	1534
	782-2	9	16	0,25	0,05	0,20	71	0,26	4,88	2,34	2323
	898-1	14	12	0,05	0,01	0,04	71	0,16	4,49	2,18	1845
3	898-2	16	11	0,04	0,01	0,03	73	0,18	4,34	1,98	1518
	898-3	13	12	0,03	0,00	0,30	61	0,11	3,70	1,69	1510
	1489-1	11	16	0,06	0,00	0,06	73	0,58	5,72	2,66	2023
4	1489-2	10	19	0,05	0,03	0,02	78	0,26	6,53	2,88	2185
	1489-3	7	17	0,02	0,01	0,01	45	0,20	4,40	2,18	1013
	1548-1	17	10	0,06	0,00	0,06	60	0,20	4,10	2,09	1620
5	1548-2	14	20	0,10	0,03	0,07	80	0,23	5,96	3,06	2138
	1548-3	10	20	0,04	0,02	0,02	57	0,26	5,07	2,47	1859
	1530-1	13	20	0,00	0,00	0,00	36	0,25	4,42	2,2	1126
6	1530-2	10	14	0,04	0,03	0,01	74	0,27	4,06	2,11	1307
	1530-3	10	15	0,07	0,03	0,04	55	0,25	4,40	2,19	1521
7	Langkawi-1	5	38	4,02	1,71	2,31	63	0,84	5,32	2,58	2286
	Langkawi-3	6	17	7,47	5,75	1,72	67	0,40	4,13	2,13	2217

Real-time PCR results confirmed the spread of the leptospires into the three main organs of the guinea pigs, i.e. lungs, liver and kidney based on detection of LipL32. In the results below, it can be concluded that the presence of leptospires was detected at the highest level in all three tissues (lungs, kidney and liver) of guinea pigs infected with strain Langkawi with cycle threshold value (Ct value) of <30 compared to other strains, indicating higher concentration of leptospires burdern. This was followed by strain 782 with a slightly lower level of Ct value in all three tissues. In animals infected with strains 898 and 1489, *Leptospira* spp. was detected in kidney and liver only, while in those infected with strain 1548 and 1530d, *Leptospira* spp. was detected only in the liver (Table A5).

Table A5. Results of real-time PCR of lung, kidney and liver from one control and six infected guinea pig groups of three.

	real time PCR Ct value						
Tissues	**Control**	**Langkawi**	**782**	**898**	**1489**	**1548**	**1530d**
lung	ND	21.27	34.65	ND	ND	ND	ND
kidney	ND	28.47	31.41	36.27	38.35	ND	ND
liver	ND	20.65	29.30	28.40	35.68	37.84	38.26

ND, not detected
Ct, cycle threshold

On histopathology findings, strain Langkawi inoculated guinea pigs moderate to severe pulmonary haemorrhage, whereas strain 782 group suffered moderate pulmonary haemorrhage. Strain 898 and strain 1489 both showed equal severity whereas strain 1548 and strain 1530d had the mild intra-alveolar haemorrhage in the lungs. Figure A5 showed the microscopic view of lung sections of guinea pigs infected with all six strains and control. In the kidneys, interstitial nephritis / pyelitis and or tubular degeneration was noted mainly in strain 782, with moderate changes seen in two out of three guinea pigs. Strains Langkawi and 898 had same intensity of mild renal histology changes in all three guinea pigs. Table A6 summarised histopathology findings in six groups of guinea pigs infected with six strains of *L. interrogans*.

Table A6. Summary of histopathology findings in one control and six infected groups of guinea pigs Strain Langkawi guinea pigs showed moderate to severe intra-alveolar haemorrhage in the lungs, with mild interstitial pyelitis and tubular degeneration in the kidneys. Strain 1530d showed mild intra-alveolar haemorrhage in the lungs.

Organ	Guinea Pig no	Control	Langkawi	782	898	1489	1548	1530d
Lung*	I	0	+++	++	0	0	+	+
	II	0	++	++	0	0	0	0
	III	0	+++	++	++	++	+	0
Kidney**	I	0	+	++	+	0	0	0
	II	0	+	0	+	0	0	0
	III	0	+	++	+	0	0	0
Liver***	I	0	0	0	0	0	0	0
	II	0	0	0	0	0	0	0
	III	0	0	0	0	0	0	0

* multifocal intra-alveolar haemorrhage
** interstitial nephritis and /or pyelitis and/or tubular degeneration
*** hepatocyte necrosis and regeneration

\+ mild , ++ moderate, +++ severe

Densiometric measurement of haemorrhage area in histologic images of lung showed that Langkawi featured the greatest extent of haemorrhage while strain 1530^{d} featured the least. The difference in haemorrhage region between the strains was significant with ANOVA $P<0.05$ (Figure A4).

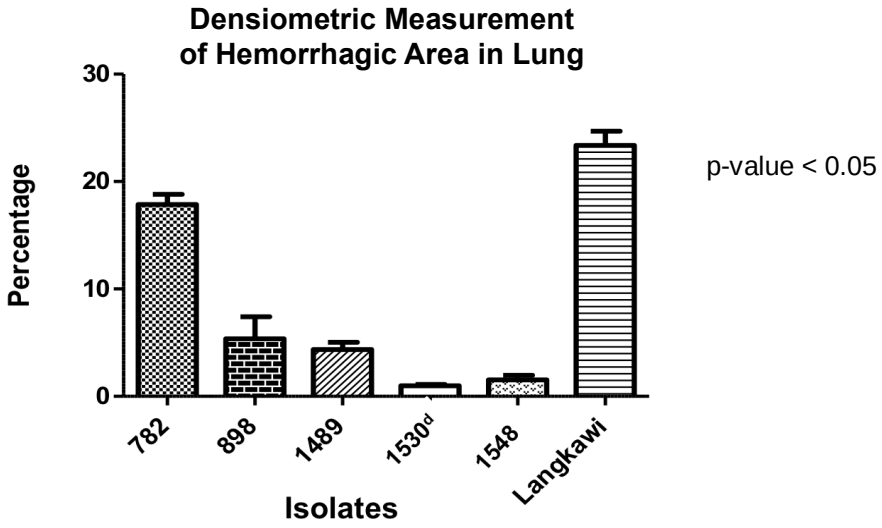


Figure A4. Histogram of densitometric measurement of intra-alveolar haemorrhage regions in lungs of infected guinea pigs. Lungs infected with strain Langkawi showed almost 25% total haemorrhage region, while those infected with strain 1530d had the least total haemorrhage region of around 2%. Across-group differences were significant at P <0.05 (ANOVA).

From the clinical point of view, patients infected with the Langkawi strain had the most complicated sequelae of pulmonary haemorrhage, which was followed by strain 782. These findings correlated well with results of the animal virulence test, i.e. strain Langkawi caused the most severe complications, followed by strain 782.

On the other hand, the patients infected with strain 898 and strain 1530^{d} were the two patients who were managed in the normal ward. Of these two patients, strain 1530^{d} did not cause any complication, whereas the patient infected with strain 898 patient developed acute kidney injury. The histopathology examination of the guinea pig infected with strain 1530^{d} showed very minimal involvement of the lungs, liver and kidney. Thus, we can conclude that clinical and animal survival test results correlated well in that the same strains were identified as the most virulent (Langkawi) and least virulent (1530d).

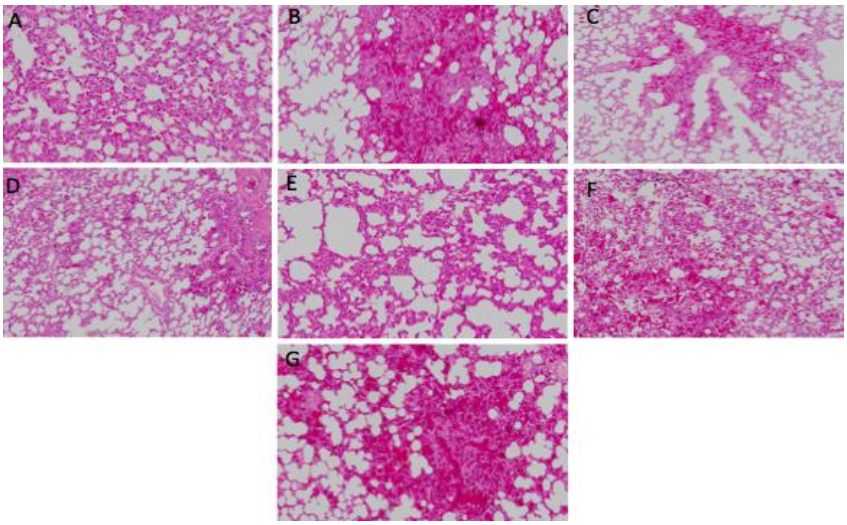


Figure A5. Microscopic 40x H&E stained lung sections from one control and six infected groups of guinea pig. (A) control: moderate lymphohistiocytic interstitial pneumonia, (B) strain 782: moderate multifocal intra-alveolar haemorrhage, (C) strain 898: moderate multifocal intra-alveolar hemorrhage; moderate multifocal lymphohistiocytic interstitial pneumonia, (D) strain 1489: moderate multifocal intra-alveolar hemorrhage with mild multifocal lymphohistiocytic interstitial pneumonia, (E) strain 1530d: mild multifocal intra-alveolar hemorrhage; mild multifocal lymphohistiocytic interstitial pneumonia, (F) strain 1548: mild multifocal intra-alveolar hemorrhage; mild multifocal lymphohistiocytic interstitial pneumonia, (G) strain Langkawi: severe multifocal intra-alveolar hemorrhage and alveolar edema.

Another animal study involving three groups of three guinea pigs, i.e. infection with strain Langkawi or strain 1530d group and control group, was then conducted in order to validate the differences between these two extreme strains in terms of virulence. Sera of all guinea pigs were sampled on day five after inoculation and analyzed for haemotological and biochemistry parameters. Infection with strain Langkawi showed slightly higher white cell count and lower platelet count compared to control and strain 1530d. Both strain Langkawi and 1530d groups caused increased liver enzymes, but only strain Langkawi caused increased total bilirubin. Infection with strain Langkawi caused raised serum creatinine, urea nitrogen and low total protein and albumin, suggestive of acute kidney injury. Serum lactate dehydrogenase (LDH) and creatinine kinase (CK) were higher in 1530d infection. (Table A7).

Table A7. Hematology and biochemistry tests results of guinea pig groups of three i.e. control, the most and the least virulence *L. interrogans* strains. There is no standard normal threshold for each biochemistry test. Thus, each result should be interpreted relative to the control group.

	Control 1	Control 2	Control 3	1530d_1	1530d_2	1530d_3	Langkawi_1	Langkawi_2	Langkawi_3
WBC	4,8	4,3	4,4	4,5	5,9	3,8	6,3	5,5	5,9
RBC	4,44	4,49	4,99	5,28	4,71	4,73	5,08	4,79	4,66
Hemoglobin	7,89	7,7	8,2	8,57	8,07	8,2	8,26	8,45	7,89
Hemotocryte	0,42	0,41	0,44	0,46	0,44	0,43	0,43	0,44	0,41
MCV	94	92	88	86	93	91	84	92	87
MCH	1,78	1,71	1,65	1,62	1,71	1,73	1,63	1,76	1,7
MCHC	18,9	18,8	18,6	18,8	18,5	19,1	19,3	19,1	19,4
Platelets	335	200	180	204	224	168	195	100	200
Eosynophils	0,14	0,14	0,12	0,03	0,03	-	0,02	0,02	0,01
Segmented neutrophils	0,4	0,51	0,42	0,54	0,44	0,55	0,44	0,4	0,61
Lymphocutes	0,45	0,35	0,45	0,4	0,53	0,43	0,49	0,53	0,34
Monocytes	-	-	-	0,03	-	0,02	0,04	0,02	0,03
AST	49	43	68	285	152	85	130	77	128
ALT	28	32	31	71	57	49	23	31	50
ALP	54	63	73	132	148	72	80	61	67
Glucose	209	210	144	190	223	190	175	268	225
Creatinine	0,6	0,5	0,6	0,6	0,6	0,5	0,9	0,5	1,2
Urea nitrogen	61	40	53	63	55	54	104	46	137
Total Protein	42	42	44	46	43	48	51	0,2	45
Total bilirubin	0,3	0,3	0,2	0,3	0,2	0,3	1,4	29	5
Albumin	30	31	31	33	31	33	33	12	28
GGT	14	9	12	10	11	12	16	10,9	11
Calcium	10,2	9,4	10	10	9,8	11,4	11	10,9	12,4
Phosphorus	6,2	4,4	7,2	8,4	7,4	8,2	10,3	6,6	10,1
Magnesium	2,1	2,3	2,3	2,6	2,4	2,9	2,8	2,1	3,9
Cholesterol	46	39	48	55	37	75	71	59	49
LDH	396	285	516	1051	767	460	411	447	482
Creatine kinase	1915	1249	1799	5691	3879	2685	676	2212	1236
Triglyceryde	68	64	43	75	56	107	128	177	153
Sodium	134,4	133,9	133,9	134,1	133,3	134,3	126	133,9	120,1
Potassium	4	4	4,4	4	4,2	3,9	4,5	4	4,5
Chloride	96	100	96,2	93	93,6	91,2	84,5	98,7	77
Globulin	12	11	13	13	12	15	18	20	17

.

4.3 Typing & Phylogeny Analysis

4.3.1 *Serotyping*

Serogroup determination was done with the MAT, using a panel of 43 polyclonal rabbit anti-*Leptospira* reference antisera, representative of 24 pathogenic and five saprophytic serogroups. The results are presented in Table A8. According to Wagenaar et al., 2005, strain Langkawi was determined to be serogroup Icterohaemorrhagiae. Strains 898 and 1530d were also identified to be from serogroup Icterohaemorrhagiae, strains 1489 and 1548 were identified to be from serogorup Bataviae, and strain 782 was identified to be from serogroup Canicola.

Table A8. Summary of serogroup typing by microscopic agglutination test with polyclonal reference anti-sera.

Strain	Serogroup typing with reference polyclonal anti sera	Titre
782	Serogroup Ballum, serovar Kenya, strain Njenga	1;80
	Serogroup Canicola, serovar Canicola, strain Hond Utrecht IV	1;1280
	Serogroup Icterohaemorrhagiae, serovar Copenhageni, strain Wijnberg	1;40
	Serogroup Javanica, serovar Poi, strain Poi	1;80
898	Serogroup Icterohaemorrhagiae, serovar Copenhageni, strain M20	1;10240
	Serogroup Icterohaemorrhagiae, serovar Icterohaemorrhagiae, strain RGA	1;10240
	Serogroup Sarmin, serovar Weaveri, strain CZ 390	1;1280
1489	Serogroup Bataviae, serovar Bataviae, strain Swart	1;2560
	Serogroup Djasiman, serovar Djasiman, strain Djasiman	1;80
1548	Serogroup Bataviae, serovar Bataviae, strain Swart	1;2560
	Serogroup Tarassovi, serovar Rama, strain 316	1;1520
	Serogroup Tarassovi, serovar Tarassovi, strain Perepelitsin	1;160
1530d	Serogroup Icterohaemorrhagiae, serovar Copenhageni, strain M20	1;10240
	Serogroup Icterohaemorrhagiae, serovar Icterohaemorrhagiae, strain RGA	1;320
	Serogroup Pyrogenes, serovar Pyrogenes, strain Salinem	1;80
	Serogroup Sarmin, serovar Weaveri, strain CZ 390	1;640

Subsequent typing of the isolate at the level of serovars was done by MAT using a panel of 18 mAbs, characteristically agglutinating serovars of the serogroup Icterohaemorrhagiae for strain 898 and 1530d, serogroup Bataviae for strain 1489 and 1548 and serogroup Canicola for strain 782. The results showed that strain 898 belonged to serovar Copenhageni (serogroup Icterohaemorrhagiae), strain 1489 belonged to serovar Paidjan (serogroup Bataviae), strain 1548 belonged to serovar Bataviae (serogroup Bataviae) and strain 1530d belonged to serovar Yeonchon (serogroup Icterohaemorrhagiae). The Langkawi strain had been serotyped in 2004 at Leptospirosis Reference Centre, Amsterdam, which was referred to as Lai-like strain (Wagenaar et al., 2005). The results of serotyping are shown in Table A9.

Table A9. Serovar typing was performed by microagglutination tests of isolates with sets of monoclonal antibodies (mAb) corresponding to specific serogroups. Five strains could be assigned to serovars but results for strain 782 were inconclusive.

	No	782*	898	1489	1548	1530d
Panel		Canicola	Icterohaemorrhagiae	Bataviae	Bataviae	Icterohaemorrhagiae
Monoclonal antibody	1	F152C1; 80	F20C3-1; 10240	F129C2-3; 20480	F129C2-3; 20480	F20C3-1; 10
	2	F152C2; 10	F20C4-3; 20480	F129C3-3; 5120	F129C3-3; 10240	F20C4-3; 10
	3	F152C5; 10	F52C1-4; 80	F129C4-1; 640	F129C4-1; 320	F52C1-4; 160
	4	F152C7; 5120	F52C2-4; 10	F129C6-1; 1280	F129C6-1; 20480	F52C2-4; 10
	5	F152C8; 640	F70C4-5; 2560	F129C7-1; 320	F129C7-1; 20480	F70C4-5; 1280
	6	F152C10; 20480	F70C13-5; 1280	F129C9-5; 10	F129C9-5; 5120	F70C13-5; 2560
	7	F152C11; 2560	F70C20-4; 2560	F129C15-3; 10	F129C15-3; 5120	F70C20-4; 640
	8	F152C13; 10	F70C24-20; 2560	F129C18-1; 10	F129C18-1; 20480	F70C24-20; 10
	9	F152C14; 2560	F70C26-3; 1280	F129C19-3; 20480	F129C19-3; 20480	F70C26-3; 1280
	10	F152C17; 10	F82C1-3; 10	F129C20-5; 20480	F129C20-5; 20480	F82C1-3; 1280
	11	F152C18; 10	F82C2-3; 10	F129C24-3; 20480	F129C24-3; 20480	F82C2-3; 40
	12		F82C7-3; 10	F129C25-1; 2560	F129C25-1; 20480	F82C7-3; 10
	13		F82C8-4; 10	F129C26-1; 10	F129C26-1; 1280	F82C8-7; 5120
	14		F89C3-3; 320			F89C3-3; 5120
	15		F89C12-6; 640			F89C12-6; 5120
Homologous Serovar		determined by CAAT	serovar Copenhageni; serovar Icterohaemorrhagiae	serovar Paidjan, strain Paidjan	serovar Bataviae, strain Swart	serovar Yeonchon, strain HM 3

* Serovar typing is inconclusive with MAT

Strain 898 is homologous to serovar Copenhageni or serovar Icterohaemorrhagiae, strain 1489 is homologous to serovar Paidjan, strain 1548 is homologous to serovar Bataviae and strain 1530d is homologous to serovar Yeonchon. The result of MAT with mAbs was not conclusive with strain 782. Thus, a CAAT was performed on strain 782. The test was started by assessing the results of cross agglutination between rabbit antiserum of serovar from serogroup Canicola with the antigen of interest, i.e. strain 782. Antigen-antiserum combination with heterogenous agglutination (50%) were later subjected to absortion tests using antiserum and antigens (live and dead) respective serovars from Canicola serogroup. The serovar with maximum absorption (least agglutination) was considered as the homologous serovar. Here, serovar Bindjei showed the highest similarity to strain 782 (Table A10).

Table A10. Summary of CAAT performed on strain 782.

Antiserum	Antigen	Cross-agglutination (%)	CAAT Live Antigen (%)	CAAT Dead Antigen (%)
Bafani	782	6.25	Nil	Nil
Benjamini	782	3.13	Nil	Nil
Malaya H6	782	< 3.13	Nil	Nil
Galtoni	782	6.25	Nil	Nil
Kuwait	782	3.13	Nil	Nil
Vleermus qoc	782	6.25	Nil	Nil
Qunjian	782	3.13	Nil	Nil
Hond Utrecht	782	50	25	25
Yonsis	782	50	25	25
Portlandvere	782	50	25	50
Sumneri	782	50	25	25
Kamituga	782	12? 62?	Nil	Nil
782	Kamituga	12? 62?	100	50
Broomi patane	782	50	0.2	0.2
782	Broomi patane	50	6.3	6.3
782	Broomi patane	50	12.5	12.5
Broomi patane	782	50	3.1	3.1
782	Broomi patane	50	12.5	12.5
782	Broomi patane	50	12.5	12.5
Bindjei	782	50	0.8	< 0.4
782	Bindjei	50	6.3	0.4
782	Bindjei	50	3.1	6.3
Bindjei	782	50	3.1	1.6
782	Bindjei	50	1.6	1.6

4.3.2 *In silico DNA-DNA Hybridization*

In order to confirm the species type of each strain, *in-silico* DNA-DNA hybridization test (DDH) was performed in which the type strain ATCC 43642T strain was used as reference. Genomes that score more than 70% similarity in DDH belonged to the same species with respect to the reference. This was the case for all strains test, thus, all strains are of the same species. The most similar strain to the reference genome, i.e. *L. interrogans* ATCC 43642T is 898 with a digital DDH estimate of about 98.4%. The most distant strain to *L. interrogans* ATCC 43642T was 1548 with 91.3% homology. The other strains DDH ranged from 93 to 93.4% (Table A11)

Table A11. *In silico* DNA-DNA hybridization of six *L. interrogans* Isolates.

Query genome	Reference genome	DDH	Model C.I.	Distance	Prob. DDH >= 70%	G+C difference
*L_interrogans*_ATCC43642	Langkawi	93	[91.1 - 94.6%]	0,0088	96.71	0,21
*L_interrogans*_ATCC43642	782	92,8	[90.8 - 94.3%]	0,0091	96.64	0,19
*L. interrogans*_ ATCC43642	898	98.4	[97.7 - 98.9%]	0.0025	97.95	0.03
*L_interrogans*_ATCC43642	1489	92,2	[90.2 - 93.9%]	0,0097	96.49	0,13
*L_interrogans*_ATCC43642	1548	91,3	[89.2 - 93.1%]	0,0107	96.22	0,15
*L_interrogans*_ATCC43642	1530	93,4	[91.5 - 94.9%]	0,0084	96.8	0,01

4.3.3 *16S rRNA typing*

16S rRNA sequences are highly conserved between different species of bacteria typing was then performed in order to differentiate strains at species-specific level. 16S rRNA typing was then performed and the result showed that all six strains of *L. interrogans* were highly similar (99% nucleotide identity) to *L. interrogans* ATCC 43642T, i.e. the type strain of *L. interrogans* which is *L. interrogans* serovar Icterohaemorrhagiae strain RGA. The findings of conventional 16S rRNA typing in agreed with DDH (Figure A6).

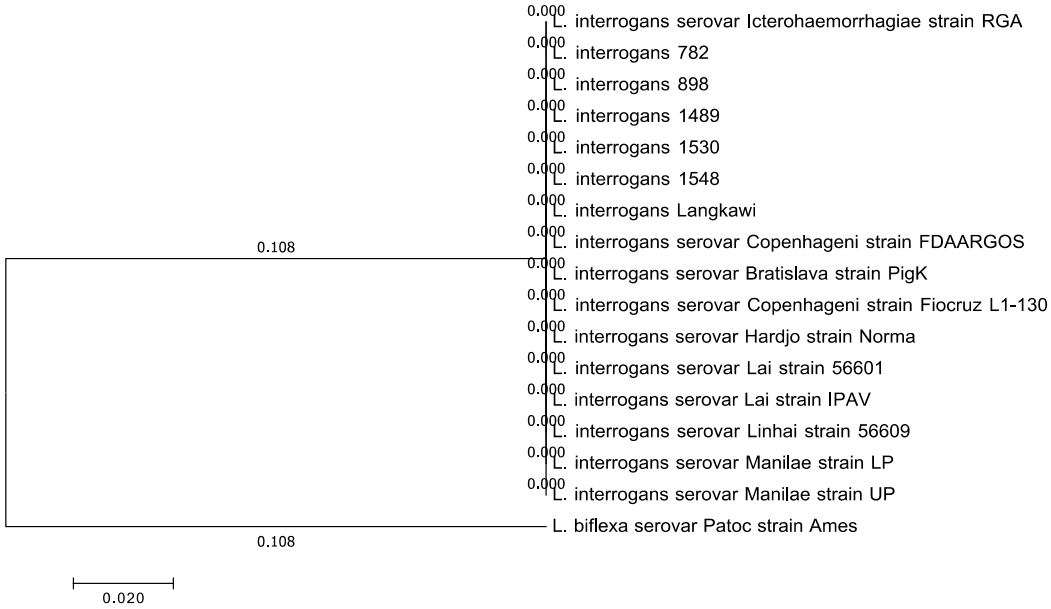


Figure A6. Molecular phylogenetic analysis of 16 strains of *Leptospira* spp. based on 16S rRNA using maximum likelihood algorithm. All eleven strains of *L. interrogans* showed almost 0.000 distance which suggested homology of almost 100%. *L. biflexia* strains showed a 10.8% difference in its 16S rRNA compared to *L. interrogans*.

4.3.4 *Multilocus sequence typing*

MLST was performed for phylogenetic analysis of six strains based on allelic variation at seven locus od conserved housekeeping genes is catalogued, and a sequence type or lineage is assigned. Application of *Leptospira* MLST Scheme 1 (Boonsilp et al., 2013) of the six genomes revealed three novel alleles and four novel ST. Sequence typing (ST) of strain 898 was identified to be ST17 which is homologous to that of *L. interrogans* serovar Copenhageni Fiocruz strain L1-130 and *L. interrogans* serovar Icterohaemorrhagiae strain RGA. Strain 1548 is identified to be ST50 which is homologous to that of *L. interrogans* serovar Bataviae strain Losbanos LT101-89. Four novel ST being assigned to the other four strains i.e. strain Langkawi, 1530d, 782 and 1489 belonging to ST236, ST237, ST240 and ST241 respectively. Strain Langkawi possesses a new allele (i.e. 71) for the *pntA* locus and strain 1530d possess new alleles at loci *pfkB* (i.e. 76) and *caiB* (i.e. 58) (Table A12).

Table A12. *Leptospira* MLST (Scheme 1) results for the six *L. interrogans* strains.

	MLST Scheme 1								Homologous
Isolate	**glmU**	**pntA**	**sucA**	**tpiA**	**pfkB**	**mreA**	**caiB**	**ST**	**Serovar/Strain**
Langkawi	1	71^	1	13	34	3	6	236*	NR
782	1	13	2	2	6	10	13	240*	NR
898	1	1	2	2	10	4	8	17	Copenhageni Fiocruz L1-130/ Icterohaemorrhagiae RGA
1489	1	1	1	4	18	5	6	241*	NR
1548	6	8	2	2	9	7	5	50	Bataviae Losbanos LT101-69
1530[d]	1	1	1	1	76^	2	58^	237*	NR

^ new allele
* novel ST
NR, not relevant

4.3.5 *Core Genome Multilocus Sequence Typing*

Core genome multilocus sequence typing (cgMLST) is a high resolution, intra-species, genome-wide, gene by gene typing. cgMLST is performed by assigning specific alleles to a predefined set of core genes, i.e., genes present in all strains of a given bacterial species (Schurch et al., 2018).

Here we use a newly derived ad-hoc cgMLST scheme typing as an attempt to characterize our six strains of *L. interrogans* strains in higher resolution and

subsequently provide insights into the relationship amongst strains circulating in Malaysia. An ad-hoc scheme of *L. interrogans* was generated using whole genome sequences of 23 *L. interrogans* strains amongst eight strains from NCBI and 15 *L. interrogans* strains that have been newly sequenced, in order to have enough WGS *L. interrogans* strains to construct the ad-hoc cgMLST scheme (Table 1).

Table A13. 23 strains of *L. interrogans* used to generate an ad-hoc cgMLST scheme for *L. interrogans.* Eight strains were previously sequenced and downloaded from NCBI GeneBank. The other 15 strains newly sequenced draft genomes.

No.	Species	Serovar	Strain	Source
1	*L.interrogans*	Lai	56601	NCBI
2	*L.interrogans*	Lai	IPAV	NCBI
3	*L.interrogans*	Hardjo	Norma	NCBI
4	*L.interrogans*	Manilae	NIID-HP	NCBI
5	*L.interrogans*	Manilae	NIID-LP	NCBI
6	*L.interrogans*	Copenhageni	Fiocruz L1-130	NCBI
7	*L.interrogans*	Linhai	56609	NCBI
8	*L.interrogans*	Bratislava	PigK151	NCBI
9	*L.interrogans*	Abramis	Abraham	New WGS^
10	*L.interrogans*	Australis	Ballico	New WGS^
11	*L.interrogans*	Autumnalis	Akiyami A	New WGS^
12	*L.interrogans*	Bataviae	Swart	New WGS^
13	*L.interrogans*	Biggis	Biggs	New WGS^
14	*L.interrogans*	Birkini	Birkin	New WGS^
15	*L.interrogans*	Djasiman	Djasiman	New WGS^
16	*L.interrogans*	Evansi	267-1348	New WGS^
17	*L.interrogans*	Gurungi	Gurung	New WGS^
18	*L.interrogans*	Haemolytica	Sejroe	New WGS^
19	*L.interrogans*	Icterohaemorhagiae	RGA	New WGS^
20	*L.interrogans*	Muelleri	RM2	New WGS^
21	*L.interrogans*	Pomona	Pomona	New WGS^
22	*L.interrogans*	Ricardi	Richardson	New WGS^
23	*L.interrogans*	Smithi	Smith	New WGS^

The studied six *L. interrogans* strains were typed using the new ad-hoc scheme. A total number of 2726 targets were defined for cgMLST (2711286 bases), 808 targets were used as accessory targets (670935 bases) and 149 targets were discarded. The reference genome was NC_004342.2, 4338762 bases, 3390 genes of *L. interrogans* serovar Lai str. 56601 chromosome I, complete sequence.; NC_004343.2, 359372 bases, 293 genes of *L. interrogans* serovar Lai str. 56601 chromosome II, complete sequence. Core genome MLST Genome coverage was 57.7% of the reference genome bases covered by cgMLST targets. A phylogenetic tree of 29 strains of *L. interrogans* constructed using neighbor-joining algorithm

based on the core genome was constructed using the ad-hoc cgMLST scheme and is displayed in Fig 1.

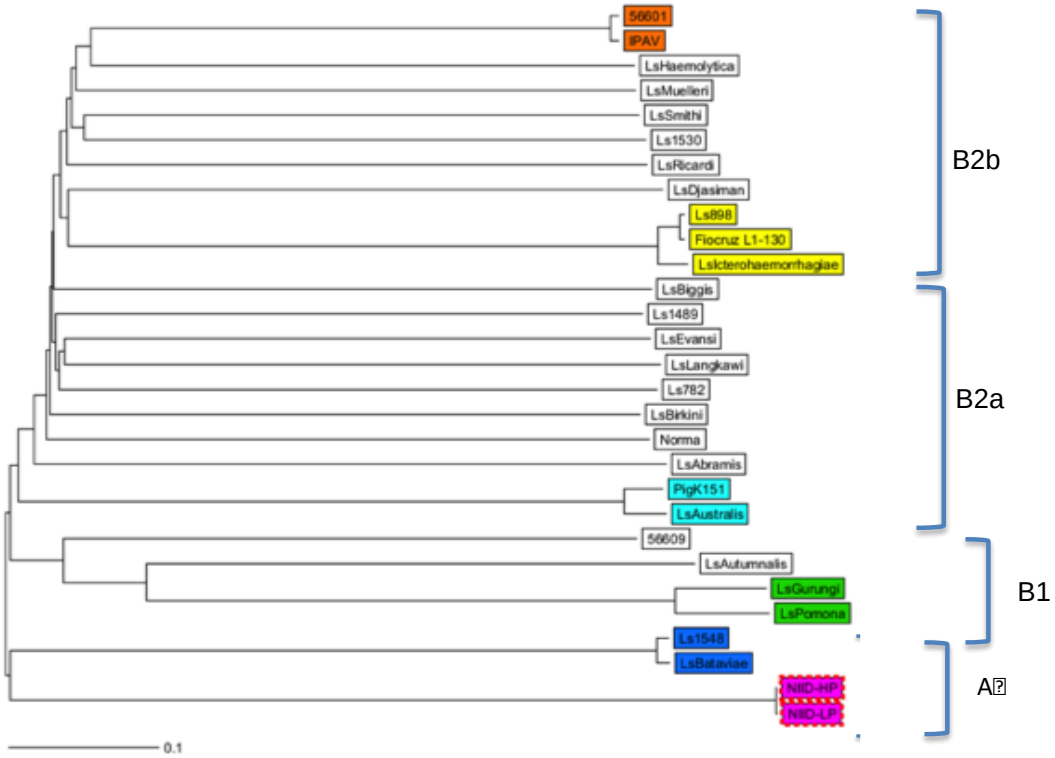


Figure A7. Phylogenetic tree based on cgMLST of *L. interrogans* using neighbour joining algorithm. There are 2726 loci in core genome with coverage of 57.7% of the reference genome. In this figure, the topology placement of strains was compared to ST of MLST Scheme 1.

The phylogenetic tree shows there are two major lineages of *L. interrogans* represented by Clade A and B. Clades A consists of four strains i.e. NIID-HP, NIID-LP, 1548 and Swart (serovar Bataviae). Clade B1, B2 are nested clades, and Clade B2 composed of two clusters a and b. Clade B1 consisted of four strains i.e. Pomona, Gurungi, Akiyami A (serovar Autumnalis) and 56609 (serovar Linhai). Clade B2a consisted of ten strains i.e. PigK, Ballico (serovar Australis), Abramis, Norma, Birkini, 782, Langkawi, Evansi, 1489 and Biggis. The second clusters (Clade B2b) consisted of eleven strains i.e. Fiocruz L1-130, RGA (serovar Icterohaemorrhagiae) 898, Djasiman, Richard (Ricardi), 1530, Smithi, Muelleri, Haemolytica, IPAV and 56601 (serovar Lai).

Four out of six *L. interrogans* case study have a sister taxa i.e. 1489 with Biggs; Langkawi with Evansi; strain 898 and Fiocruz L1-130, 1548 and Swart (serovar Bataviae).

4.3.6 SNP genotyping

In addition, another whole genome sequence typing was applied based on SNP calling. Herewith, a phylogenomic tree was constructed based on SNP calling using the parSNP software (Treangen et, al. 2014). The same set of *L. interrogans* strains were used as in cgMLST. Within parSNP, trees are constructed using a maximum likelihood algorithm based on the SNP calling within a core genome determined by 'sequence clusters'.

By using parSNP, a total number of 2841 sequence clusters were created, with an average number of MUMs per cluster was 16 (average cluster length was 1307 bp). The total coverage of core genome among all strains was 77.5%. The phylogenetic tree of 29 strains of *L. interrogans* based on SNPs is displayed in Fig 2.

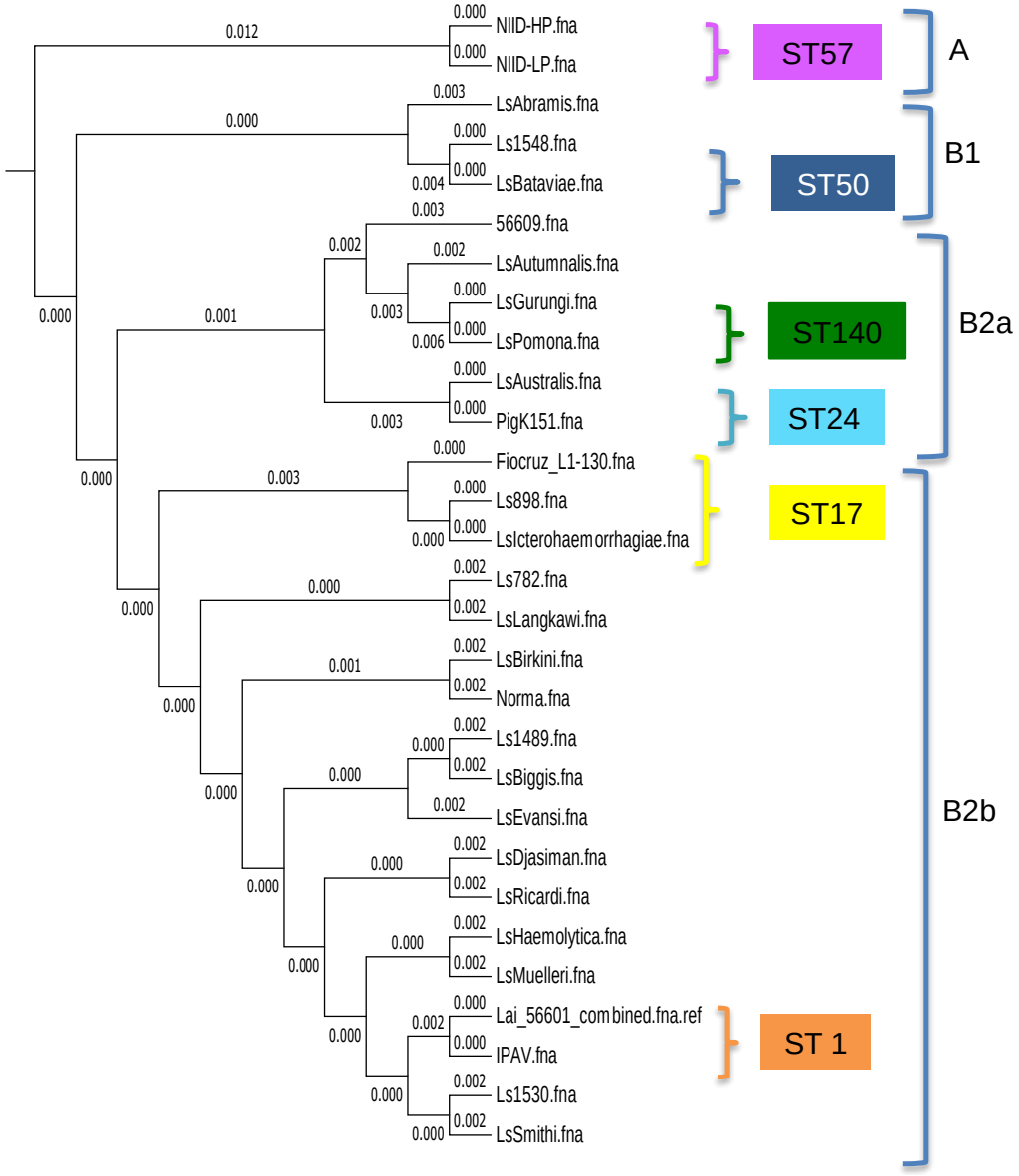


Figure A8. Phylogenetic tree of *L. interrogans* based on SNPs calling by ParSNP using maximum likelihood algorithm.

The phylogenetic tree of SNP calling shows that there are two major lineages of *L. interrogans* represented by Clade A and B. Clades A consists of only two strains i.e. NIID-HP, NIID-LP (serovar Manilae). Clade B consists of two nested clades i.e. Clade B1 and B2. Clade B1 constituted of strain Abramis, 1548 and Swart (Bataviae). Clade B2 are divided into two clusters a and b; Clade B2a constituted of strain 56609, Akiyami A (serovar Autumnalis), Gurung (serovar Gurungi), Pomona, Ballico (serovar Australis) and PigK151; while Clade B2b constituted of Fiocruz L1-130, 898, Icterohaemorrhagiae; 782 and Langkawi; Biggs (serovar Biggis), 1489,

Birkin (serovar Birkini) and Norma; Clade 6 constituted of Djasiman, Ricardi, Haemolytica, Muelleri, Lai 56601, IPAV, 1530, Smithi and Evansi.

4.3.7 Comparison of MLST, cgMLST and SNP genotyping

From all molecular typing methods used in this study MLST is the simplest as the strains are not completely sequenced and only use six or seven housekeeping genes are used. However, there are limitations as it comes to discrimination of bacteria on strain level. For instance, strains with similar STs could not be differentiated the strains further as seen in strain 898, Fiocruz L1-130 and serovar Icterohaemorrhagiae which belong to ST50. The discriminations of strains can be carried out easily by cgMLST and ParSNP.

From our results we can also conclude that MLST is highly congruent to cgMLST and SNP calling as strains with similar ST i.e. ST1, ST17, ST24, ST50, ST57 and ST140 are being grouped as sister elements in the latter two analyses. Interestingly, in SNP calling, the distance of these similar ST sister groups are within 0.0-0.5% of the branch length. In cgMLST, the similar ST sister groups have differences in 2 genes (between serovar Manilae NIID-HP and NIID-LP) to 323 genes differences as seen between strain Gurungi to Pomona.

There are two major clades of *L. interrogans* in both typing methods however, SNP calling have a more distinctive grouping of Clade B by having six nested clades compared to two in cgMLST and SNP genotyping. Topology placements of the strains within the two main clades are also different. The strains that are placed in Clade B1 cgMLST belonged to Clades B3 and B6 in SNP calling; while strains in Clade B2 of cgMLST, belonged to Clade B2, B4, B5 and exceptionally one Clade B6 strain (i.e. Evansi).

Among the six isolates, the three isolates strains 898, 1530d and 1548 were consistently placed with their sister taxa (Fiocruz L1-130 and Icterohaemorrhagiae for 898, Smithi for 1530d and Bataviae for 1548). Both strain 782 and Langkawi were closely placed cgMLST, however in SNP calling they are sister taxa. This is likewise for strain 1489 and Biggis, which are sister taxa in SNP genotyping but only closely placed in cgMLST. Interestingly, strain 898 is placed closest to strain Fiocruz L1-130 in cgMLST, with genes difference of 15 in cgMLST, branch length difference of 0.1%

in SNP genotyping. However, in SNP calling 898 was placed closes to strain Icterohemorrhagiae, as the length difference is 0.00% in SNP calling and genes difference of 95 in cgMLST. Based on epidemiological prevalence data in Malaysia, the latter is more valid as strain Fiocruz-L1-130 has never being isolated in Malaysia, unlike strain Icterohaemorrhagiae, which is one of the most circulating *L. interrogans* strains in Malaysia (Amran, 2011).

Both trees were drawn to a legend scale (ie. 1.5 cm and 2 cm), by different calculator with branch lengths representing differences of around 0.2% for ParSNP and 10% of *L. interrogans* core genome in cgMLST. This means the cluster containing strains Manilae NIID-HP and NIID-LP differs from the root around 1.2% for parSNP and 42.5% for cgMLST. The cluster containing strains 898, Fiocruz 130-L1-130 and Icterohaemorrhagiae differs from the roots around 0.466% on the nucleotide level for parSNP and 35% on the gene level in cgMLST.

Since all 29 strains in the analysis were sporadic cases and not related epidemiologically, a cluster-typing (CT) threshold was not calculated with both cgMLST and SNP calling. Although strain Langkawi and strain 782 were topologically the closest among the six strains (which may suggest that there were some related profiles between strain Langkawi and strain 782), none of the six *L. interrogans* strains were identified to be from a similar source.

4.4 Genomics Descriptions

4.4.1 *General parameters*

All genomes possessed two chromosomes. The first larger chromosome ranged from 4.28 to 4.56 Mb while the second still conserved, but smaller chromosome ranged from 3.5 to 3.95 kb. The number of plasmids ranged from 0 to 7, the number of contigs ranged from 2 to 10. Genome size of each ranged from 4.6 to 5.3 Mb with 3.6 k to 4.4 k coding sequences. All strains had two 23S, two 16S and one 5S rRNA gene. As previously described (Fukunaga et al., 1989), rRNA genes in *L. interrogans* are not organised in operons, as in most other bacteria, but are scattered across the first chromosome. All the strains have one *rrf* gene, two *rrl* genes, and two *rrs* genes

encoding 5S, 23S, and 16S rRNA, respectively, as reported in Nascimento et al., 2004. Miscellaneous RNA was variable in number. It was the highest in number in strain 782 but was not found in strain 1530d. The general parameters of the six *L. interrogans* genomes are summarized in Table A13. The largest genome belonged to strain 1489 with 5.3Mb while the smallest genome belonged to strain 898. The high number of plasmids in strain 1489 increased the size of the genome in comparison to strain 898, which does not contain a plasmid and thus has the smallest genome (Table A14).

Table A14. General parameters of five complete genomes and a draft genome of *L. interrogans* strains. Strain Langkawi, 782, 898, 1489 and 1548 are complete genomes, strain 1530 was not fully resolved.

	Langkawi	**782**	**898**	**1489**	**1548**	**1530d***
Size (bp)	4,885,316	5,308,174	4,630,592	5,349,767	4,880,112	5,013,967
Chromosome 1 (Mb)	4.37	4.56	4.28	4.43	4.35	4.32
Chromosome 2 (Mb)	0.39	0.35	0.35	0.35	0.39	0.35
No of Plasmid	2	5	0	7	2	≥5
No of Contigs	4	7	2	9	4	9
CDS	3916	4212	3618	4270	3857	3943
tmRNA	1	1	1	1	1	1
repeat region	1	2	1	2	3	2
Gene	4048	4476	3716	4516	4000	4087
tRNA	37	37	37	37	37	37
23S	2	2	2	2	2	2
16S	2	2	2	2	2	2
5S	1	1	1	1	1	1
Signal peptide	132	129	124	133	131	125
Miscellaneous RNA	5	11	5	10	5	0

d, draft genome

* not fully resolved

Only 1 to 24% of the sequences assigned to plasmids could be identified in GenBank by blastn, which suggested that the remaining regions of the plasmid were novel or only distantly related to currently known genomes. Most common plasmid sequences were from *L. interrogans* serovar Canicola strain Gui44 plasmid pGui1, which was found in two copies in strain 1489 and strain 1530d. The results were summarized in Table A15.

Table A15 Plasmids identification and distribution in six isolates of *L. interrogans*. Color code refers to same plasmids.

Strain	Unitig	Size (bp)	Region	Blastn	Identification	E-value
Langkawi	13	395410	22699-41023	*Leptospira interrogans serovar Lai strain 56601 plasmid Laicp*	99	0.0
	15	590330	22565-28568	*Leptospira interrogans serovar Linhai strain 56609 plasmid lcp3*	99	0.0
782	3	101385	1026883-1033920	*Leptospira interrogans serovar Linhai strain 56609 chromosome 1*	97	0.0
	4	90872	1827270-1839364	*Leptospira interrogans serovar Hardjo strain Norma chromosome 1*	97	0.0
	5	74118	29277-41023	*Leptospira interrogans serovar Lai strain 56601 plasmid Laicp*	99	0.0
	6	53819	32956-41866	*Leptospira interrogans serovar Linhai strain 56609 plasmid lcp3*	90	0.0
	103	77059	44165-52502	*Leptospira interrogans serovar Linhai strain 56609 plasmid lcp1*	99	0.0
1489	2	140106	3395202-3405478	*Leptospira interrogans serovar Hardjo strain Prajitno chromosome 1*	96	0.0
	3	83109	64228-74981	*Leptospira interrogans serovar Canicola strain Gui44 plasmid pGui1*	99	0.0
	8	76580	920797-928587	*Leptospira interrogans serovar Bratislava strain PigK151 chromosome 1*	86	0.0
	12	93306	38618-52502	*Leptospira interrogans serovar Linhai strain 56609 plasmid lcp1*	99	0.0
	14	39186	3135926-3145302	*Leptospira interrogans serovar Hardjo strain Norma chromosome 1*	98	0.0
	369	62980	18199-24647	*Leptospira interrogans serovar Linhai strain 56609 plasmid lcp2*	99	0.0
	372	69692	48710-59098	*Leptospira interrogans serovar Canicola strain Gui44 plasmid pGui1*	97	0.0
1548	3	80435	4106-14814	*Leptospira interrogans serovar Manilae strain UP-MMC-NIID-LP plasmid pLIMLP1*	91	0.0
	4	51265	1024339-1034087	*Leptospira interrogans serovar Linhai strain 56609 chromosome 1*	95	0.0
1530[d]	58	91738	1– 9103	*Leptospira interrogans serovar Linhai strain 56609 chromosome 1*	97	0.0
	59	82644	3169915-3188778	*Leptospira interrogans serovar Hardjo strain Norma chromosome 1*	99	0.0
	60	52032	1024362-1034085	*Leptospira interrogans serovar Linhai strain 56609 chromosome 1*	97	0.0
	61	54623	64228-74956	*Leptospira interrogans serovar Canicola strain Gui44 plasmid pGui1*	99	0.0
	62	52974	32956-41860	*Leptospira mayottensis 200901116 plasmid lcp2_L200901116*	90	0.0
	103	2169	353896-356065	*Leptospira interrogans serovar Lai strain IPAV chromosome 2*	100	0.0

Unitig, contig number

E-value, refers to alignment scores to NCBI GenBank

The functional annotation of proteins in the plasmids of five *L. interrogans* strains (without strain 898) was then investigated using Clusters of Orthologous Groups (COG). Protein of unknown function were the main proteins found in each plasmid in five *L. interrogans* strains except for plasmid of unitig 3 in strain 782 and plasmid of unitig 103 in strain 1530 which mainly consisted of replication, recombination and repair proteins; and transcription and signal transduction mechanisms proteins in their respective plasmid gene (Figure A9).

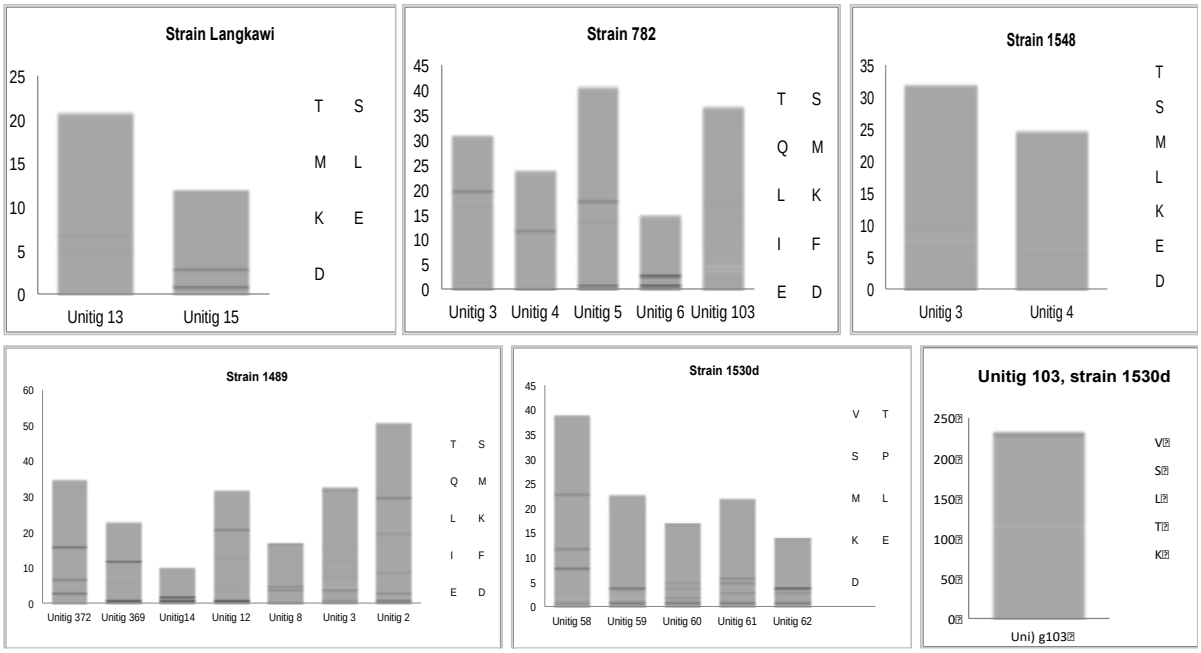


Figure A9. Functional annotation of proteins in the plasmids of five *L. interrogans* strains (without strain 898). Some more explanations needed so that the reader knows what was done. But the results belong into the main text. D: cell cycle control, cell division, chromosome partitioning; E: amino acid transport and metabolism; F: nucleotide transport and metabolism; I: lipid transport and metabolism; K: transcription; L: replication, recombination and repair; M: cell wall, membrane, envelope biogenesis, P: inorganic ion transport ad metabolism; Q: secondary metabolites biosynthesis, transport and catabolism; R: general function prediction only; S: function unknown; T: signal transduction mechanism; V: defense mechanism.

4.4.3 *Protein orthology*

Among the six strains, the total number CDS which are present in all genomes (core genome) is 3271 whereas CDS that are unique to each strain (strain-specific CDS) ranged from 50 to 352. The number of shared CDS for each strain was as follows: Langkawi, 477; strain 782, 694; strain 898, 297; strain 1489, 647; strain 1548, 383; and strain 1530d, 530. The intermediate CDS overlapping between each genome are not shown (Figure A10).

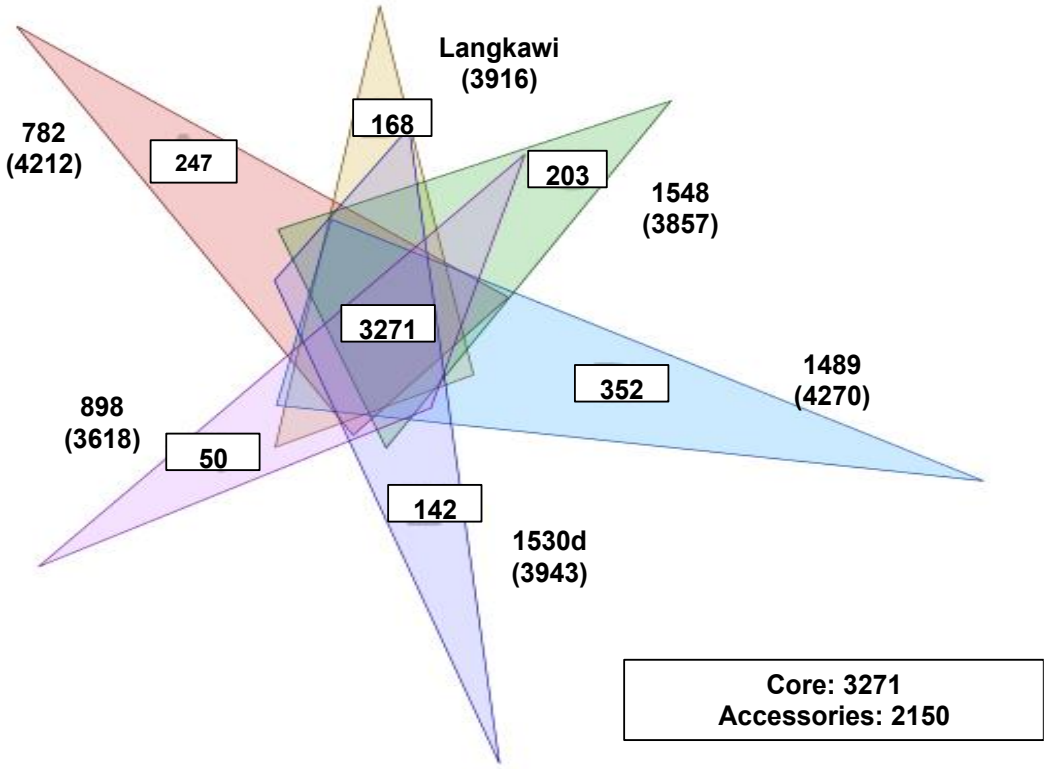
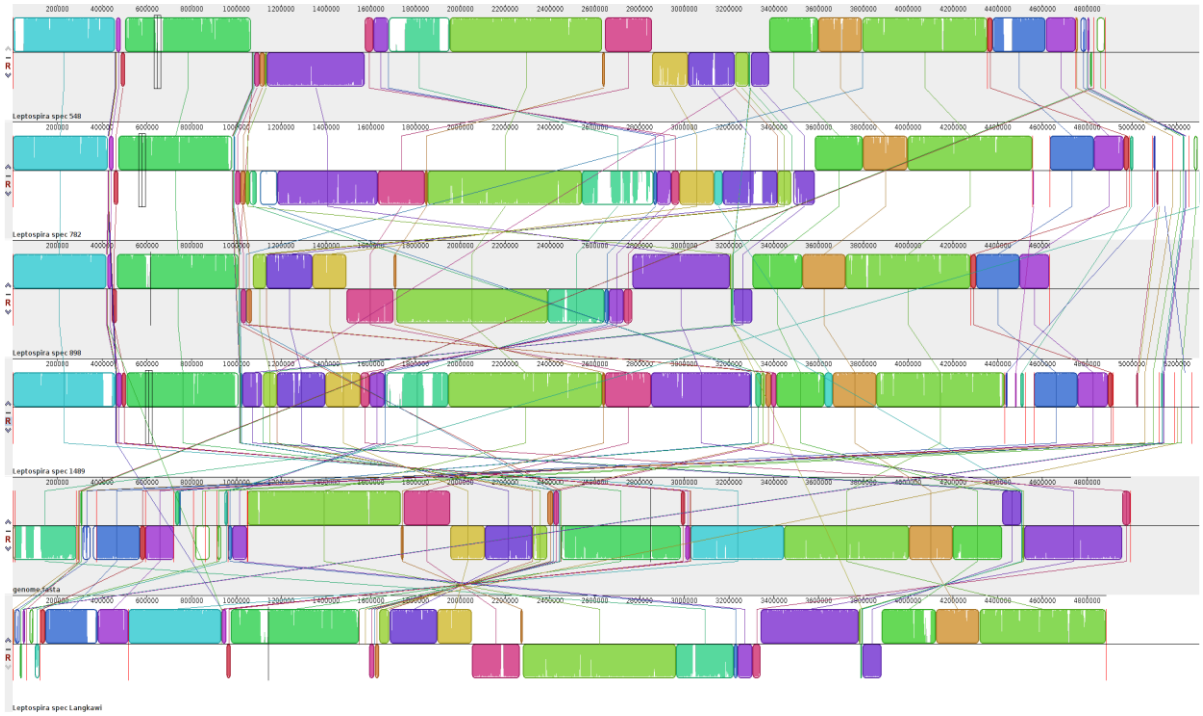


Figure A10. Venn diagram of six *L. interrogans* isolates displaying the core CDS and unique CDS of each strain.

4.4.4 *Protein synteny*

Conserved and syntenic regions were visualized using MAUVE in terms of large collinear blocks (LCBs) among the six *L. interrogans* strains represented by different colored boxes (Figure A11). Strain-specific regions i.e. genomic islands (GI) hereby were represented either by white regions within LCBs or by spaces in between. Although the sizes of LCBs are more or less conserved across the strains, they were substantially differently ordered, suggesting different genome structures and rearrangements. Nevertheless, all six strains of *L. interrogans* were widely collinear / syntenic.



Conserved regions, colored lines / boxes; specific regions (genomic island), white boxes

Figure A11. Protein synteny of six *L. interrogans* strains displayed in large collinear blocks . The LCB are represented mainly by conserved regions, as represented by coloured lines/boxes, which were organised in different order. From this LCB, it can be concluded that all six *L. interrogans* strains were widely collinear/synthenic.

4.4.5 *Prophages*

Numerous bacterial virulence factors are known to contribute to the development of infectious diseases in eukaryotes, some of which are transferred by mobile genetic elements such as prophages (Brüssow et al 2004). Furthermore, phages may also act as significant agents of bacterial evolution by enabling genetic exchange and recombination, which may enhance or lessen the virulence potential of a bacterial pathogen.

We thus performed in silico phage finding by submitting the nucleic acid sequences to the online PHASTER tool. The results showed that the range of phages found in the genome strains was estimated between 1 to 10. Strain Langkawi and strain 782 have the most phages in their genome with 7 and 10, respectively, whereas the two strains from the Bataviae serogroup only possess 1 phage each. 898 and 1530d have four and five phages, respectively.

The number of phages in the genome might have driven genome evolution. Nevertheless, all phages found are incomplete and therefore are most likely remnants of earlier phage infections which were probably eradicated by a fully functional CRISPR-Cas system (compare section 4.4.6). Some of the remnant phage genes even seem to be frameshifted and therefore represent pseudogenes. A summary of identified prophages is displayed in Table A16.

Table A16. Prophages identification and distribution in six *L. interrogans* strains.

Strain	Total	Unitig	Region	Region Length	Completeness	Score	No. Total Proteins	Region Position	Phage Identification	Detail
Langkawi	7	3	1	8.6Kb	incomplete	30	12	237357-246004	PHAGE_Salisa_1_	Terminase,tail
		40	1	9.7Kb	incomplete	60	11	1047304-1057067	PHAGE_Lister_B054_	head,capsid
		40	2	10.5Kb	incomplete	30	14	1058927-1069497	PHAGE_Deep_s_D6E_	Plate,tail
		40	3	9.4Kb	incomplete	10	10	1426160-1435647	PHAGE_Bacill_SP_15_	-
		40	4	11.4Kb	incomplete	10	9	2621637-2633036	PHAGE_Sphing_PAU_	-
		40	5	6.9Kb	incomplete	30	7	2709383-2716331	PHAGE_Synech_S_CAM7_	lysis
		40	6	6.2Kb	incomplete	40	8	3211720-3217923	PHAGE_Pseudo_LMA2_	Tail, plate
782	10	0	1	23Kb	incomplete	30	13	1155136-1178139	PHAGE_Bacill_PfEFR_5_	transposase
		0	2	7.2Kb	incomplete	50	9	1226026-1233243	PHAGE_Pseudo_LMA2_	Plate,tail
		0	3	5Kb	incomplete	50	8	1628022-1633023	PHAGE_Bacill_SP_15_	transposase
		0	4	8.6Kb	incomplete	30	11	2641470-2650117	PHAGE_Salisa_1_	Terminase,tail
		0	5	11.3Kb	incomplete	10	9	2772751-2784143	PHAGE_Sphing_PAU	-
		0	6	6.9Kb	incomplete	30	7	2867731-2874679	PHAGE_Synech_S_CAM7_	lysis
		0	7	9.4Kb	incomplete	10	10	3077867-3087353	PHAGE_Synech_ACG_2014b_	-
		0	8	9.9Kb	incomplete	30	13	3161600-3171587	PHAGE_Psychr_pOW20_A	tail
		0	9	8.6Kb	incomplete	30	12	3337482-3346129	PHAGE_Salisa_1_	Terminase,tail
		6	10	5.8Kb	incomplete	20	11	30880-36691	PHAGE_Mannhe_vB_MhS_535AP2	tail
1489	1	0	1	28.3Kb	incomplete	60	33	1730220-1758602	PHAGE_Sphing_PAU_	transposaae
898	4	0	1	9.4Kb	incomplete	10	10	1386689-1396161	PHAGE_Bacill_SP_15_	-
		0	2	11.4Kb	incomplete	10	10	2563962-2575364	PHAGE_Sphing_PAU_	-
		0	3	11.6Kb	incomplete	30	11	2647794-2659399	PHAGE_Synech_S_CAM7_	lysis
		0	4	6.2Kb	incomplete	40	8	3153143-3159346	PHAGE_Pseudo_LMA2_	Tail,plate
1548	1	0	1	8.4Kb	incomplete	30	9	3037136-3045545	PHAGE_Acidia_virus_	transposase
1530d	5	0		6.2Kb	incomplete	40	8	308140-314343	PHAGE_Pseudo_LMA2_	-
		0	2	7Kb	incomplete	30	7	2538121-2545213	PHAGE_Lactob_Lj965_	Tail,plate
		0	3	9.4Kb	incomplete	10	11	2813059-2822530	PHAGE_Bacill_SP_15_	transposase
		0	4	11.3Kb	incomplete	10	9	4024631-4036023	PHAGE_Sphing_PAU_	-
		0	5	11.6Kb	incomplete	30	11	4109976-4121581	PHAGE_Synech_S_CAM7_	lysis

4.4.6 *Clustered Regularly Interspaced Short Palindromic Repeats Analysis*

CRISPR is a short form of 'Clustered Regularly Interspaced Short Palindromic Repeats'. It represents the prokaryotic adaptive immune system that confers resistance to foreign genetic elements such as prophages (Marraffini et al., 2010, Mojica et al., 2009). CRISPRs consist of a succession of 24–50 bp long direct repeats or 'repeats' separated by similarly sized unique sequences called spacers. CRISPR spacers are derived from phage DNA and extrachromosomal DNA such as plasmids from viruses that previously tried to attack the bacteria (Pourcel et al., 2005, Mojica et al., 2005, Bolotin et al., 2005).

According to Fouts et al., 2016, only infectious *Leptospira spp*. possess CRISPR system in their genome. Confirming this result we found between 10 to 16 loci with 1 to 13 spacers in the CRISPR arrays in each of the six *L. interrogans* strains. However, the number of confirmed CRISPR arrays was the highest in strain 1548 i.e. total of 6 CRISPR arrays while the least are in strain 782, 898 and 1489 had 2 each (Table A17).

Table A17. Identification and distribution of CRIPSR arrays in six *L. interrogans* strains.

Strains	Total	Contig	Direct Repeat Consensus	No of Spacers	Length (bp)	Location
Langkawi	**16**	Unitig_40	CGCCTAACGGCATCAAAGTTATATTCAG	3	246	1166186-1166432
		Unitig_40	CGCCTAACGGCATCAAAGTTATATTCAG	7	524	1168294-1168818
		Unitig_40	TTCCTAAAGAAATAGGGAATTTAAAAAA	4	304	1492306-1492610
		Unitig_40	GTAGTTTTCCGATTTCTTTTGGAA	6	443	2936207-2936650
782	**16**	Unitig_0	TTCCTAAAGAAATCGGAAAACTAC	6	442	1512608-1513050
		Unitig_0	TGAATATAACTTTGATGCCGTTAGGCGTTGAGCA	13	965	3415949-3416914
898	**11**	Unitig_0	GTGCTCAACGCCTAACGGCATCAAAGTTATATTCAG	3	253	1133696-1133949
		Unitig_0	TTCCTAAAGAAATAGGGAATTTAAAAAA	4	304	1451034-1451338
1489	**15**	Unitig_0	GTGCTCAACGCCTAACGGCATCAAAGTTATATTCAG	4	323	1180651-1180974
		Unitig_0	GTGCTCAACGCCTAACGGCATCAAAGTTATATTCAG	12	901	1181131-1182032
1548	**13**	Unitig_0	TTCCTAAAGAAATCGGAAAACTAC	6	443	1460500-1460943
		Unitig_0	CTGAATATAACTTTGATGCCGGTAGGCGTTGAGCAC	12	900	3224282-3225182
		Unitig_0	CTGAATATAACTTTGATGCCGGTAGGCGTTGAGCAC	3	255	3225339-3225594
		Unitig_0	CTGAATATAACTTTGATGCCGGTAGGCGTTGAGCAC	5	396	3226341-3226737
		Unitig_0	CTGAATATAACTTTGATGCCGGTAGGCGTTGAGCAC	5	395	3226895-3227290
		Unitig_0	CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC	3	248	3234838-3235086
1530d	**10**	Unitig_0	GTGCTCAACGCCTAACGGCATCAAAGTTATATTCAG	7	540	2554190-2554730
		Unitig_0	GTGCTCAACGCCTAACGGCATCAAAGTTATATTCAG	3	250	2554888-2555138
		Unitig_0	GTGCTCAACGCCTAACGGCATCAAAGTTATATTCAG	9	683	2555295-2555978

The most commonly found spacer element originated from plasmid *lcp1* from *L. interrogans* serovar Linhai, followed by plasmid *lcp3* from the same serovar. Some other spacers were found to be of hypothetical proteins or unknown genetic

elements (not shown). Strain 1530d displayed the fewest CRISPRs with no identifiable genetic elements in the spacers. One of the spacers in strain 898 was 'flagellar type III secretion system protein FlhB' which was most likely due to gene duplication (Table A18).

Table A18. Identification and distribution of CRISPR spacers In six *L. interrogans* strans.

Strains	CRISPR	Frequecncy	CRSIPR	Blastn (spacer)
Langkawi	Repeat region	1	CGCCTAACGGCATCAAAGTTATATTCAG	*Leptospira interrogans* serovar Linhai str.56609"plasmid *lcp3*,
	Spacer	2	AAAATCGTTGTTGCGAAATGCGGATCGTTGTTCAGGTGCTCAA	product = "putative bacteriophage tail fiber protein"
	Repeat region	4	CGCCTAACGGCATCAAAGTTATATTCAG	*Leptospira interrogans* serovar Linhai str.56609"plasmid *lcp1*,
	Spacer	1	CTTGAATCGAAGTGTTAGAAATTGAAAATTTTTTCGGTGCTCAA	product="hypothetical protein"
	Repeat region	6	CGCCTAACGGCATCAAAGTTATATTCAG	*Leptospira interrogans* serovar Linhai str.56609"plasmid *lcp3*,
	Spacer	5	TTCGTCAAAGACGAGAGAGTATCTCAAGAGAAGAGTGCTCAA	product="Helix-turn-helix domain containing protein"
782	Repeat region	12	CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC	*Leptospira interrogans* serovar Linhai str.56609"plasmid *lcp1*,
	Spacer	1	ATCATGCGAATCCGGTGGGAACGTTGCTGGTTTAT	product="putative bacteriophage tail fiber protein"
	Repeat region	12	CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC	*Leptospira interrogans* serovar Linhai str.56609"plasmid *lcp3*,
	Spacer	2	ATCCAAATTGTCCTTTCCCCATTGCCTGTAGGGAT	product="hypothetical protein"
	Repeat region	12	CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC	*Leptospira interrogans* serovar Linhai str.56609"plasmid *lcp3*,
	Spacer	3	CAATTGTGATCGGAATCGGTTTCCAAAATTTAAG	product="putative phage replication related protein"
	Repeat region	12	CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC	*Leptospira interrogans* serovar Linhai str.56609"plasmid *lcp1*,
	Spacer	5	GATGAAACAAAAGTCCCTCGAATTCAAGTGAGGCT	product="hypothetical protein"
	Repeat region	12	CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC	*Leptospira interrogans* serovar Linhai str.56609"plasmid *lcp1*,
	Spacer	7	AGGTCCACCCCTTCCTGAGTAGACACAGGCAACGGA	product="hypothetical protein"
	Repeat region	12	CTGAATATAACTTTGATGCCGTTAGGCGTTGAGCAC	*Leptospira interrogans* serovar Linhai str.56609"plasmid *lcp1*,
	Spacer	8	ACAAAAGCAGCATATCTAGACTTATCCAAAATAGAT	product="hypothetical protein"
898	Repeat region	6	GAGTTCCCACAATTTACACGAGATC	*Leptospira interrogans* serovar Copenhageni FDAARGOS_203
	Spacer	1	AAAAATCTTATGGGAGTTCCCACAATTTACACGAGATCAAAAATCTTATGG	product="flagellar type III secretion system protein FlhB"
1489	Repeat region	4	GTGCTCAACGCCTAACGGCATCAAAGTTATATTCAG	*Leptospira interrogans* serovar Hardjo-prajitno str. Hardjo
	Spacer	3	TCTTCCGAAACGGAAACACGCCAAGGGTCTATAATTT	product="plasmid replication initiation factor"
	Repeat region	4	GTGCTCAACGCCTAACGGCATCAAAGTTATATTCAG	*Leptospira interrogans* serovar Linhai str.56609"plasmid *lcp1*,
	Spacer	8	AGGGGCGCTTTTCCGGGTCCTCATTGTTCTATTTT	product="hypothetical protein"
	Repeat region	4	GTGCTCAACGCCTAACGGCATCAAAGTTATATTCAG	*Leptospira interrogans* serovar Linhai str.56609"plasmid *lcp1*,
	Spacer	9	TTCCAGCATTGGCGGGGAACCCTCAAGCATTGCAAT	product="hypothetical protein"
	Repeat region	4	GTGCTCAACGCCTAACGGCATCAAAGTTATATTCAG	*Leptospira interrogans* serovar Linhai str.56609"plasmid *lcp1*,
	Spacer	12	GCGCTGTCGTTTGCGGGCAAGCCGTAAGACCGCAT	product="hypothetical protein"
1548	Repeat region	8	CTGAATATAACTTTGATGCCGGTAGGCGTTGAGCAC	*Leptospira interrogans* serovar Linhai str.56609"plasmid *lcp3*,
	Spacer	1	AGTTACTCAAAAATACTCAATTAGAGGAGTATTGGGGA	product="hypothetical protein"
1530d	Repeat region	3	GTGCTCAACGCCTAACGGCATCAAGTTATATTCAG	ND
	Spacer	1	CACTTTCTCTTGAATAAGAGAAAGAAAC	ND

ND, not detected

4.4.7 *Insertion Sequences*

Pathogenic leptospires possess more than 20 insertion sequences (IS) their genome and ISs have been reported to be a major contributor to genomic evolution and diversification (Fouts et al., 2016). Table A19 summarise the number and type of IS in each strain. It was noted that strain 1489 has the highest number of IS followed by strain 1548 and 1530d. The majority of IS belonged to ISLin1 and were located in chromosome 1. Less than 30 IS were detected in strain 898 and strain 782. As only IS with a score of equal to e-value 0.0 were taken as significant, the IS with larger e-value scores were disregard.

Table A19. Identification and distribution of IS in six *L. interrogans* strains. ISLin1 is the most common IS in all 6 strains followed by ISLin2. The highest IS was located in strain 1489 followed by 1548 (Bataviae serogroups).

	Contig	ISLin1	ISLin2	IS500A	IS500B	IS1501	IS1502	IS1533
Langkawi	40	25	8	ND	ND	ND	1	4
	3	7	ND	ND	ND	ND	ND	ND
	2	ND	ND	ND	ND	ND	ND	ND
	15	ND	ND	ND	ND	ND	ND	ND
782	0	6	1	ND	ND	ND	ND	ND
	2	ND	ND	ND	ND	ND	ND	ND
	3	ND	ND	ND	1	ND	ND	ND
	103	ND	ND	ND	ND	1	ND	ND
	5	ND	1	ND	ND	ND	ND	ND
898	0	ND	1	8	ND	7	ND	ND
	4	4	ND	ND	ND	ND	ND	ND
1489	0	91	ND	ND	1	6	2	24
	1	7	ND	ND	ND	ND	ND	ND
	2	ND	ND	ND	ND	1	ND	ND
	3	ND	3	ND	ND	ND	ND	ND
	12	ND	ND	ND	1	1	ND	ND
	372	ND	ND	1	ND	1	ND	ND
	369	ND	ND	ND	ND	2	ND	ND
1548	0	65	9	ND	ND	7	2	2
	2	5	1	ND	ND	ND	ND	ND
	3	ND	ND	ND	ND	1	ND	ND
	4		4	ND	ND	ND	ND	ND
1530d	0	59	8	ND	ND	4	1	3
	2	1	ND	ND	ND	1	ND	ND
	57	7	ND	ND	1	ND	ND	ND
	58	ND	ND	ND	ND	1	ND	ND
	59	ND	ND	1	ND	ND	ND	ND
	61	ND	1	ND	ND	ND	ND	ND

ND, not detected

Here, we also report IS elements which possibly transfer certain bacterial virulence factors. We found an UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase in strain Langkawi and strain 898, which encodes a lipopolysaccharide biosynthesis protein similar to the *Bordetella pertussis* Tohama I protein, flanked by a remnant transposase ISLin1 (Family IS110). Besides that, a putative GTP pyrophosphokinase RelA (ATP: GTP 3'-pyrophosphotransferase) (PPGPP synthetase I) ((P)PPGPP synthetase) (GTP diphosphokinase) [RelA] similar to the *Mycobacterium tuberculosis* H37Rv gene is embedded within IS of ISLin1 (Family IS110) type in strain 898 and strain 782. Strain 782 possesses two other virulence factors, i.e. GDP-mannose 6-dehydrogenase algD [Alginate] similar to the

Pseudomonas aeruginosa (P. aeruginosa) PAO1 gene and beta-hemolysin [β-hemolysin] from *Staphylococcus aureus* subsp. aureus COL inserted by ISLin1 (Family IS110). dTDP-glucose 46-dehydratase from *Haemophilus influenzae* Rd KW20 was inserted in strain 1489 ISLin1 (IS110). ISLin1 in strain 1548 inserts one virulence factor for alginate biosynthesis, i.e. D-alanyl-lipoteichoic acid biosynthesis protein dltB, whereas no virulence factor inserted by IS elements in strain 1530d was detected.

4.5 Comparative Genomic Analysis

4.5.1 O-antigen

4.5.1.1 Rfb locus

The O-antigen is the basis for *Leptospira* serovar identification. Serovar classification of leptospires is based on the expression of the epitopes arranged in a mosaic pattern on the surface of the lipopolysaccharide (LPS) layer. The specificity of epitopes depends on their sugar composition and orientation (Levett, 2015; Adler, 2015; Faine et al.,1999). Due to the high importance of the O-antigen in the pathogenesis of an infection, the *rfb* loci of the six *L. interrogans* strains were analysed.

The *rfb* locus of *L. interrogans* strains is located between the *marR* (transcriptional regulator) and *sdcS* (Na+/dicarboxylate symporter) genes (Fouts et al., 2016). Taking *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 as reference the *rfb* loci of all six strain were compared. The crude pattern and arrangement of gene clusters of O-antigen correlate with the serotyping results, i.e. Fiocruz, strain 898, strain 1530^{d} and strain Langkawi belong to the Icterohaemorrhagiae serogroup while strain 1489 and strain 1548 belong to the Bataviae serogroup. Strain 782 is different in arrangement and gene cluster as it belongs to the Canicola serogroup. Strain 898 and strain 1530^{d} are similarly organised (Figure A12). Strain Langkawi was similarly organised as serogroup Icterohaemorrhagiae as well, however, there are three additional genes, which are marked by green, orange and blue circles in the figure. The largest *rfb* locus is present in strain 782, which is enlarged by two large insertions (red circles). Interestingly, the gene arrangements of strain 1498 and strain 1548 are very different compared to all strains described before, while they are

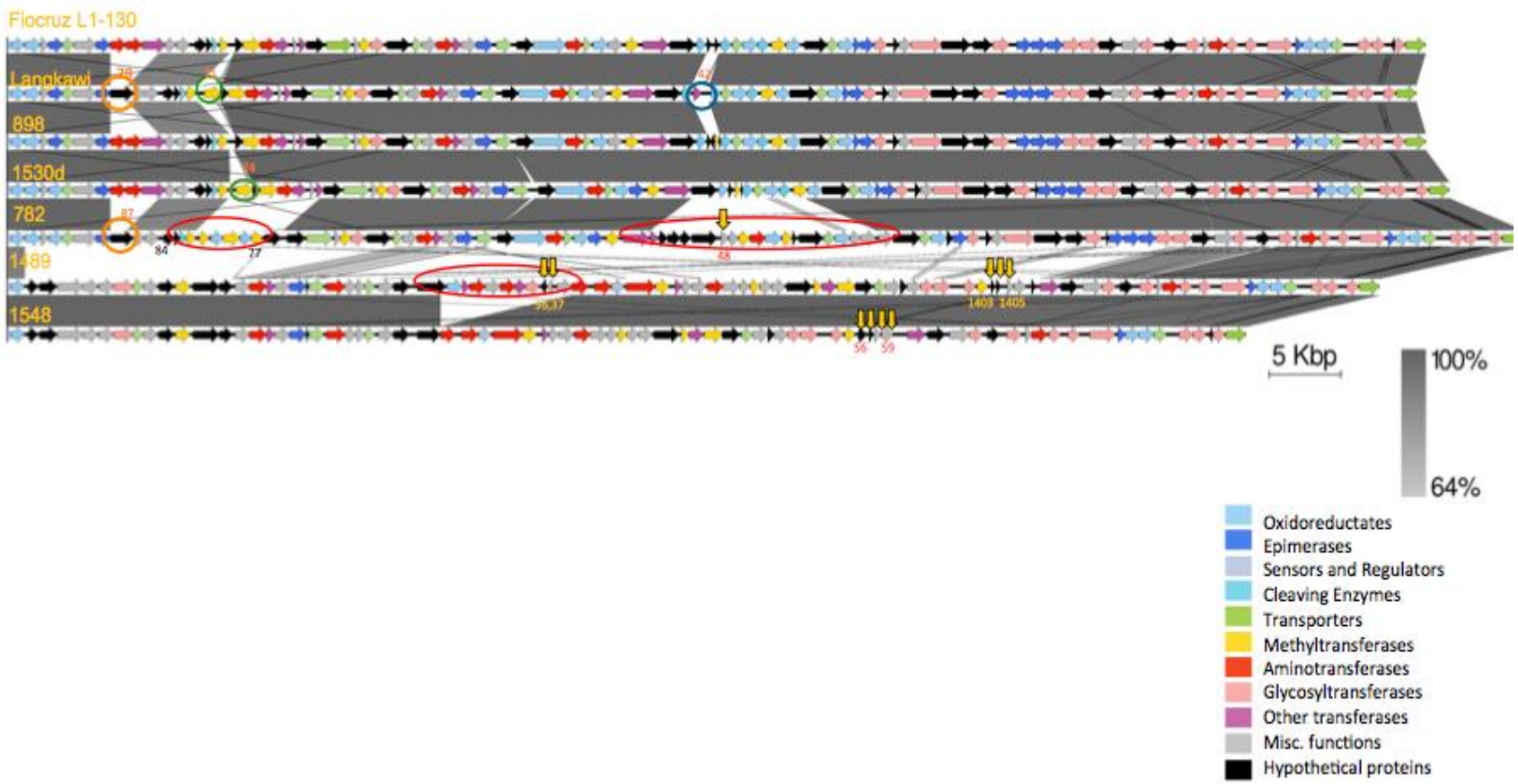


Figure A12. *Rfb* locus of six *L. interrogans* isolates with reference to *L. interrogans* serovar Copenhageni strain Fiocruz L1-130. The crude pattern and arrangement of gene clusters of the O-antigen correlate with the serotyping results. Additional gene clusters within strain 782 and 1489 are marked by red circles. Inserted transposases within two of the clusters are indicated by yellow arrows. Strain Langkawi displays a gene combination unique from 782 (orange circle), 1530 (green circle) as well as a Galactoside O-acetyltransferase (blue circle), which makes it entirely different from the other Icterohaemorrhagiae serogroup strains. Strains Fiocruz, 898 and 1530d possess similar gene arrangements.

similar in arrangement and gene content between each other. Strain 1489, however, possesses an additional gene cluster (red circle). The transposases in the *rfb* locus in strains 782 and 1489 are shown by yellow arrows and identified and described in Table A19. However, in both strains, the transposase genes are sequentially positioned, probably due to frameshift mutations. Interestingly, transposase genes are not present in the other three strains belonging to the Icterohaemorrhagiae serogroup, which leads to the hypothesis that the *rfb locus* of Icterohaemorrhagiae is possibly more stably arranged.

Table A20. Identification and distribution of transposases in gene clusters of the *rfb* locus in three *L. interrogans* strains. In strains 1489 and 1548 transposase genes are sequentially positioned, probably due to frameshift mutations.

Strain	Locustag	Blastp	Identification (%)	E-value
782	782_02405	putative helix-turn-helix domain of transposase IS66, partial [*Leptospira interrogans* str. UI 08452]	94	9.00E-77
1489	1489_01371	transposase, IS4 family [*Leptospira kirschneri* str. 2008720114]	100	1.00E-84
	1489_01372	MULTISPECIES: IS5 family transposase [*Leptospira*]	100	7.00E-80
	1489_01403	transposase, IS4 family [*Leptospira interrogans* serovar Bataviae str. L1111]	100	1.00E-112
	1489_01404	IS4 family transposase [*Leptospira kirschneri*]	98	0.0
	1489_01405	transposase DDE domain protein [*Leptospira kirschneri* str. 2008720114]	100	5.00E-60
1548	1548_01437	transposase [*Leptospira interrogans* serovar Bataviae]	100	2.00E-146
	1548_01438	ISPg7, transposase family protein [*Leptospira noguchii* str. 2007001578]	100	1.00E-53
	1548_01439	transposase, IS4 family [*Leptospira interrogans* serovar Bataviae str. L1111]	100	2.00E-112
	1548_01440	transposase, IS4-like family protein [*Leptospira interrogans* serovar Bataviae str. L1111]	100	0.0

4.5.1.2 Lipid

Although leptospiral lipopolysaccharides (LPS) resemble standard gram negative LPS chemically and immunologically, it is substantially less active in endotoxin activity. Whether this reduced biological activity is due to unique properties of leptospiral Lipid A, remains indefined (Que-Gewirth et al., 2004; Werts et al., 2001).

A study performed on *L. interrogans* lipid A showed that different acyl chains and novel phosphorylation on the position of the lipid A lead to abrogated endotoxinogenicity (Que-Gewirth et al., 2004). Furthermore, LPS from *L. interrogans* cells is unusual in that it activates TLR2 rather than TLR4. *L. interrogans* lipid A induces tumour necrosis factor with about one-tenth the potency of E. coli lipid

(Que-Gewirth et al., 2004). The lipid A biosynthetic pathway of *L. interrogans* serovar Lai involves 13 enzymes, encoded by the genes *lpxA, lpxC, lpxD1, lpxD2, lpxB1, lpxB2, lpxK, kdtA, kdsB1, kdsB2, let, kdsA* and *htrB*.

In this comparison among the available *L. interrogans* strains, the lipid A biosynthesis pathway amino acid sequences showed uniform homology except in the protein *kdsB1,* which encodes the enzyme 3-deoxy-manno-octulosonate cytidylyltransferase. Expectedly, strains which belong to Icterohaemorrhagiae showed higher sequence similarity to *kdsB1* (as the reference sequence is from a similar serogroup). According to Fouts et al., 2016, *kdsB1* and kdsB2 were only found in two species / serovar (*L. interrogans* serovar Lai and *L. inadai* serovar Lyme). Other species / serovars only had one *kdsB* gene that showed a higher level of similarity with *kdsB2.* Further studies should be performed to see if this protein can be used as a typing or functional biomarker among *L. interrogans* strains. This study shows that lipid A composition of the six *L. interrogans* strains are differentiated by *3-deoxy-manno-octulosonate cytidylyltransferase* which categorized the strains into 2: the high similarity to serovar Lai 56601 and of Icterohaemorrhagiae serogroup, i.e. strain Langkawi, strain 898 and strain 1530d (which all belonged to serogroup Icterohaemorhagiae) and low similarity, i.e. strain 782, strain 1489 and strain 1548. The sequences of the other 12 enzymes involved in lipid A biosynthesis have similarities of 99 -100% (Table A21).

Table A21. Protein homology of Lipid A proteins in six *L. interrogans* strains with reference to strain Fiocruz L1-130. The protein homology of all six *L. interrogans* strains were almost the same except in gene *kdsB1,* protein LA_1625.

Gene locus/Protein name	GenBank Accession	Protein name	Langkawi	782	1489	898	1548	1530d	Fiocruz
lpxA / LA_3949	NP_714129.1	UDP-N-acetylglucosamine acyltransferase	100%	100%	100%	100%	100%	100%	100%
lpxC / LA_2306	NP_712487.2	JDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	99%	99%	100%	99%	100%	100%	99%
lpxD1 / LA_0512	NP_710694.1	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	100%	100%	100%	100%	100%	100%	100%
lpxD2 / LA_4326	NP_714506.1	UDP-3-O-[3-hydroxymyristoyl]	99%	99%	99%	99%	99%	99%	99%
lpxB1 / LA_0786	NP_710967.1	arabinose kinase	100%	99%	100%	99%	99%	99%	99%
lpxB2 / LA_1096	NP_711277.2	lipid-A-disaccharide synthase	100%	99%	100%	100%	100%	100%	100%
lpxK / LA_3095	NP_713875.1	tetraacyldisaccharide 4'-kinase	100%	99%	99%	99%	99%	100%	99%
kdtA / LA_1477	NP_711657.1	3-deoxy-D-manno-octulosonic-acid transferase	100%	99%	100%	100%	99%	100%	100%
kdsB1 / LA_1625	NP_711806.1	3-deoxy-manno-octulosonate cytidylyltransferase	99%	53%	53%	99%	53%	99%	99%
kdsB2 / LA_3573	NP_713753.2	3-deoxy-manno-octulosonate cytidylyltransferase	100%	100%	100%	100%	100%	100%	100%
lnt / LA_4078	NP_714258.2	apolipoprotein N-acyltransferase	100%	100%	100%	100%	100%	100%	100%
kdsA / LA_2408	NP_712589.2	2-dehydro-3-deoxyphosphooctonate aldolase	100%	100%	100%	100%	100%	100%	100%
htrB / LA_4039	NP_714219.2	lipid A biosynthesis lauroyl acyltransferase	100%	100%	100%	100%	99%	100%	100%

4.5.1.3 Sialic acid

Sialic acids are 9-carbon backbone derivatives of neuraminic (Neu) and Ketodeoxynonulosonic (Kdn) acids. Sialic acids are part of a larger family of nonulosonic acid (NulO) molecules that includes pseudaminic and legionaminic acids. Sialic acids as post-translational modifications were found to be restricted to pathogenic leptospires.

Based on Fouts et al., 2016 all genes involved in biosynthesis of sialic acid are present in *L. interrogans*. In our study, we found the amino acid sequence similarity for 12 out of 13 proteins differing for strains 1489 and 1548, suggesting that both strains have slightly different protein sequence for sialic acid biosynthesis especially for N-acetylneuraminic (sialic) acid synthetase (*sas/neuB2*) and UDP-N-acetylglucosamine 2-epimerase (Table A22).

Table A22. Protein homology of sialic acids in six *L. interrogans* strains with reference to strain Fiocruz L1-130. Strains 1489 and 1548 showed general difference for 12 out of 13 proteins associated with sialic acids biosynthesis.

Protein Name	Genebank Accesion	Functions	Langkawi	782	898	1489	1548	1530d
LA_1605	NP_711786.1	CMP-N-acetylneuraminic acid synthetase (neuA1)	100%	100%	100%	59%	59%	99%
LA_1606	NP_711787.1	NAD dependent epimerase dehydratase family protein (rfbB3)	100%	100%	100%	93%	93%	99%
LA_1607	NP_711788.1	pyridoxal phosphate dependent aminotransferase	100%	100%	100%	79%	79%	100%
LA_1608	NP_711789.2	sugar acyltransferase, sialic acid O- acetyltransferase NeuD family (neuD)	100%	100%	100%	74%	74%	100%
LA_1609	NP_711790.1	pseudaminic acid synthetase (neuB)	100%	100%	100%	79%	79%	100%
LA_1610	NP_711791.1	UDP-N-acetylglucosamine 2-epimerase (neuC)	100%	100%	100%	78%	78%	100%
LA_1611	NP_711792.1	nucleotidyl transferase	100%	100%	100%	67%	67%	100%
LA_1612	NP_711793.1	hypotheticalprotein	100%	99%	100%	58%	58%	99%
LA_1613	NP_711794.1	N-acetylneuraminic (sialic) acid synthetase (sas/ neuB2)	99%	99%	99%	39%	39%	99%
LA_1614	NP_711795.1	pyridoxal phosphate dependent aminotransferase	100%	100%	100%	52%	52%	100%
LA_1615	NP_711796.1	CMP-N-acetylneuraminic acid synthetase (neuA2)	100%	100%	100%	53%	53%	100%
LA_1645	NP_711826.1	UDP-N- acetylglucosamine 2-epimerase	99%	99%	99%	42%	42%	99%
LA_3823	NP_714003.1	UDP-N-acetylglucosamine diphosphorylase	100%	100%	100%	100%	100%	100%

4.5.2 Molecular aspect of pathogenesis mechanisms

4.5.2.1 Motility

Motility is a major factor for virulence and pathogenesis of leptospirosis. Leptospires have hooked ends, two periplasmic flagella with polar insertions located in the periplasmic space. This ‘apparatus’ equips the leptospires for a unique ‘gliding-crawling’ (Tahara et al., 2018) motility, which enables them to pass through the host skin and tissue barriers efficiently (Nascimento et al., 2004 and Picardeau., 2017). There are more than 70 genes associated with motility and its regulation reflects its major role in pathogenesis (Nascimento et al., 2004). Any deficiencies in this main

set of motility genes may attenuate virulence. *flaA* and *flaB* proteins constitute the flagellar sheath and core, respectively. Deficiencies in *flaB* cause the pathogen to be non-motile and attenuates virulence (Picardeau et al., 2001).

Here, we identified protein homology of structural proteins associated with motility functions in leptospires among the six *L. interrogans* strains in comparison to Fiocruz L1-130. There were slight differences in the sequences of proteins encoded by LIC11846, LIC11392 and LIC11394, i.e. genes *fliN, fliF* and *fliH*. As expected, all these genes are 100% identical on the amino acid sequence level in strain 898, which is very near in relationship to the reference strains used (*L. interrogans* serovar Copenhageni strain Fiocruz L1-130). Another interesting finding was that only strain 1489 has a slight sequence homology to the reference strain in gene *flhB,* which codes for export apparatus (Table A23).

Table A23. Protein homology of structural motility function in six *L. interrogans* strains with reference to strain Fiocruz L1-130.

LIC ORFs	Gene	Accession No.	Annotation	Langkawi	782	1489	898	1548	1530d	Fiocruz
BASAL BODY										
LIC10297	*fliE*	*AAS68924.1*	Flagellar hook-basal body complex prote	100%	100%	100%	100%	100%	100%	100%
LIC10298	*flgC*	*AAS68925.1*	Flagellar basal-body rod protein	100%	100%	100%	100%	100%	100%	100%
LIC10299	*flgB*	*AAS68926.1*	Flagellar basal-body rod protein	100%	100%	100%	100%	100%	100%	100%
LIC10621	*flbD*	*AAS69242.1*	Flagellar protein	100%	100%	100%	100%	100%	100%	100%
LIC10622	*motA*	*AAS69243.1*	Chemotaxis protein	100%	100%	100%	100%	100%	100%	100%
LIC12931	*motA1*	*AAS71482.1*	Chemotaxis protein	100%	100%	100%	100%	100%	100%	100%
LIC10623	*motB*	*AAS69244.1*	Chemotaxis protein	100%	100%	100%	100%	100%	100%	100%
LIC11713	*motB2*	*AAS70302.1*	Motility protein B	100%	100%	100%	100%	100%	100%	100%
LIC10624	*fliL*	*AAS69245.1*	Flagellar protein	100%	100%	100%	100%	100%	100%	100%
LIC11187	*flgD*	*AAS69794.1*	Flagellar hook assembly scaffolding prot	99%	100%	100%	100%	99%	99%	100%
LIC11324	*flgG*	*AAS69927.1*	Flagellar basal-body rod protein	100%	100%	100%	100%	100%	100%	100%
LIC11325	*flgA*	*AAS69928.1*	Chaperone P-ring formation protein	100%	100%	100%	100%	100%	100%	100%
LIC11326	*flgH*	*AAS69929.1*	Flagellar L-ring protein	99%	100%	99%	100%	99%	100%	100%
LIC11327	*flgI*	*AAS69930.1*	Flagellar P-ring protein - hook-associate	100%	100%	100%	100%	100%	100%	100%
LIC11370	*fliN ©*	*AAS69971.1*	Flagellar motor switch protein/C ring	100%	100%	100%	100%	100%	100%	100%
LIC11371	*fliO*	*AAS69972.1*	Flagellar protein Type III secretion	99%	99%	99%	100%	99%	99%	100%
LIC11372	*fliP*	*AAS69973.1*	Flagellar protein biosynthesis	100%	100%	100%	100%	100%	100%	100%
LIC11373	*fliQ*	*AAS69974.1*	Flagellar protein	100%	100%	100%	100%	100%	100%	100%
LIC11374	*fliR*	*AAS69975.1*	Flagellar protein	99%	99%	99%	100%	99%	99%	100%
LIC11376	*flhA*	*AAS69977.1*	Flagellar protein	100%	100%	100%	100%	100%	100%	100%
LIC11377	*flhF*	*AAS69978.1*	Flagellar biosynthesis regulator	100%	100%	100%	100%	100%	100%	100%
LIC11380	*fliA*	*AAS69981.1*	RNA polymerase sigma factor	100%	100%	100%	100%	100%	100%	100%
LIC11389	*flbB*	*AAS69989.1*	Flagellar protein	100%	100%	100%	100%	100%	100%	100%
LIC11390	*fliJ*	*AAS69013.1*	Chaperone-like protein export apparatu	100%	100%	100%	100%	100%	100%	100%
LIC11391	*fliI*	*AAS69014.1*	Flagellum-specific ATP synthase/ATPase	100%	99%	99%	100%	99%	99%	100%
LIC11392	*fliF*	*AAS69015.1*	Flagellar M-ring protein	100%	98%	96%	100%	100%	98%	100%
LIC11394	*fliH*	*AAS69017.1*	Flagellar assembly protein/ATPase comp	100%	98%	100%	98%	100%	98%	98%
LIC11836	*fliM*	*AAS70422.1*	Flagellar motor switch protein/C ring	100%	100%	100%	100%	100%	100%	100%
LIC11846	*fliN*	*AAS70432.1*	Flagellar motor switch protein	92%	92%	92%	100%	92%	92%	92%
LIC10023	*fliG*	*AAS68660.1*	Flagellar motor switch protein/C ring	100%	100%	100%	100%	100%	100%	100%
LIC11393	*fliG1*	*AAS69993.1*	Flagellar motor switch protein/C ring	99%	100%	100%	100%	100%	100%	100%
LIC11900	*fliG3*	*AAS70485.1*	Flagellar motor switch protein/C ring	100%	100%	100%	100%	100%	100%	100%
LIC11071	*flgM*	*AAS69678.1*	Anti sigma factor for FliA	100%	100%	100%	100%	100%	100%	100%
LIC11375	*flhB*	*AAS69976.1*	Export apparatus	99%	99%	93%	100%	100%	99%	100%
LIC13452	*flgN*	*AAS71992.1*	Chaperone export apparatus	100%	100%	100%	100%	100%	100%	100%
LIC13449	*fliW*	*AAS71989.1*	Flagellar assembly factor	99%	99%	99%	100%	99%	99%	100%
LIC11562	*flhX*	*AAS70158.1*	FlhB-like protein Type III secretion	100%	98%	100%	100%	100%	100%	100%

(cont.)

HOOK										
LIC13450	*flgL*	*AAS71990.1*	Flagellar hook-associated protein	100%	100%	100%	100%	100%	100%	100%
LIC13451	*flgK*	*AAS71991.1*	Flagellar hook-associated protein	99%	100%	99%	100%	99%	100%	100%
LIC12035	*flhO*	*AAS70610.1*	Hook protein/FliG	100%	100%	100%	100%	100%	100%	100%
LIC11188	*flgE*	*AAS69795.1*	Flagellar hook protein	100%	100%	100%	100%	100%	100%	100%
LIC10723	*fliD*	*AAS69344.1*	Filament cap	99%	99%	100%	100%	100%	99%	100%
LIC11186	*fliK*	*AAS69793.1*	Flagellar hook-length control protein	99%	99%	99%	100%	99%	99%	100%
LIC11328	*flgJ*	*AAS69931.1*	Peptidoglycan hydrolase/rod cap	100%	100%	100%	100%	100%	100%	100%
FILAMENT										
LIC10788	*flaA*	*AAS69404.1*	Flagellar filament outer layer protein A	100%	99%	100%	100%	100%	100%	100%
LIC10787	*flaA2*	*AAS69403.1*	Flagellar filament outer layer protein A	100%	100%	100%	100%	100%	100%	100%
LIC11531	*flaB*	*AAS70127.1*	Flagellar filament core protein FlaB	100%	100%	100%	100%	100%	100%	100%
LIC11890	*flaB2*	*AAS70476.1"*	Flagellar filament core protein FlaB	100%	100%	100%	100%	100%	100%	100%
LIC11889	*flaB3*	*AAS70475.1*	Flagellar filament core protein FlaB	100%	100%	100%	100%	100%	100%	100%
LIC11532	*flaB4*	*AAS70128.1*	Flagellar filament core protein FlaB	100%	100%	100%	100%	100%	100%	100%
LIC11657	*fliS*	*AAS70250.1*	Flagellar protein (modular protein)	100%	100%	100%	100%	100%	100%	100%

4.5.2.2 Chemotaxis

Chemotaxis genes are well conserved among spirochetes, however, the chemotaxis apparatus of *L*. *interrogans* is likely much more complicated, as its genome contains approximately twice as many methyl-accepting chemotaxis proteins (MCP) as either *T*. *pallidum* or *B*. *burgdorferi* (Nascimento et al., 2004).

Chemotaxis proteins play important roles in survival and adaptation of pathogenic leptospires to a variety of environments and hosts. The summary of protein homology analysis of known leptospires chemotaxis proteins in reference to strain Fiocruz L1-130 is shown in Table A24.

Table A24. Amino acid sequences of chemotaxis associated proteins in six *L. interrogans* strains with reference to strain Fiocruz L1-130. Sequences of all six strains were nearly identical to the reference.

LIC ORFs	Gene	Accession No.	Annotation	Langkawi	782	1489	898	1548	1530d	Fiocruz
CHEMOTAXIS										
LIC10685	*mcp*	*AAS69306.1*	Methyl-accepting chemotaxis protein	100%	99%	99%	100%	99%	99%	100%
LIC11216	*mcp*	*AAS69822.1*	Methyl-accepting chemotaxis protein	100%	100%	100%	100%	100%	100%	100%
LIC11407	*mcp*	*AAS70007.1*	Methyl-accepting chemotaxis protein	99%	99%	99%	100%	99%	99%	100%
LIC11488	*cheX*	*AAS70086.1*	Chemotaxis protein	100%	100%	100%	100%	100%	100%	100%
LIC11520	*cheB3*	*AAS70116.1*	Chemotaxis response regulator protein	99%	100%	99%	100%	99%	99%	100%
LIC11521	*cheD1*	*AAS70117.1*	Chemoreceptor glutamine deamidase	100%	100%	100%	100%	100%	100%	100%
LIC11522	*cheW*	*AAS70118.1*	Chemotaxis protein	100%	100%	100%	100%	100%	100%	100%
LIC11523	*mcp*	*AAS70119.1*	Chemotaxis protein	99%	99%	99%	100%	99%	99%	100%
LIC11524	*cheA2*	*AAS70120.1*	Chemotaxis protein histidine kinase	99%	99%	99%	100%	99%	99%	100%
LIC11525	-	AAS70121.1	Anti sigma factor antagonist	100%	100%	100%	100%	100%	100%	100%
LIC11526	*cheY*	*AAS70122.1*	Chemotaxis response regulator	100%	100%	100%	100%	100%	100%	100%
LIC11691	*mcp-2*	*AAS70280.1*	Methyl-acceptingchemotaxis protein	100%	100%	100%	100%	100%	100%	100%
LIC11871	*cheR*	*AAS70457.1*	Chemotaxis protein methyltransferase	100%	100%	100%	100%	100%	100%	100%
LIC12061	*cheB2*	*AAS70632.1*	Chemotaxis protein-glutamate methylesterase	100%	100%	100%	100%	100%	100%	100%
LIC12062	*cheR1*	*AAS70633.1*	Chemotaxis protein	100%	99%	100%	100%	99%	100%	100%
LIC12240	*cheW2*	*AAS70812.1*	Chemotaxis protein	100%	100%	100%	100%	100%	100%	100%
LIC12455	*cheB1*	*AAS71021.1*	Chemotaxis response regulator protein	100%	100%	100%	100%	100%	100%	100%
LIC12456	*cheA*		Chemotaxis protein	99%	99%	100%	100%	99%	99%	100%
LIC12457	*cheW1*	*AAS71022.1*	Chemotaxis protein	100%	100%	100%	100%	100%	100%	100%
LIC12482	*mcp*	*AAS71047.1*	Methyl-acceptingchemotaxis I	100%	100%	100%	100%	100%	100%	100%
LIC12500	*mcpB*	*AAS71065.1*	Chemoreceptor transmembrane protein	100%	100%	100%	100%	100%	100%	100%
LIC12921	*mcpA*	*AAS71473.1*	Methyl-acceptingchemotaxis protein	99%	99%	99%	99%	99%	99%	99%
LIC13183	*mcp*	*AAS71728.1*	Methyl-accepting chemotaxis-like	100%	100%	100%	100%	100%	100%	100%
LIC13394	*dmcA*	*AAS71934.1*	Methyl-accepting chemotaxis protein	100%	100%	100%	100%	100%	100%	100%
LIC13401	*mcp*	*AAS71941.1*	Methyl-accepting chemotaxis protein	100%	100%	100%	100%	100%	100%	100%

4.5.2.3 Adhesion

Various specific ligands and adhesins facilitate adhesion of leptospires to eukaryotic extracellular matrix (ECM) molecules. The presence of these proteins is important to facilitate pathogenesis of infection in eukaryotes hosts. According to Fouts et al., 2016; *Lsa23, Lsa26, Lsa33, Lsa45, Lsa66, LipL32 and Mfn1* were found in all infectious leptospiral species (pathogens and intermediates) but absent in saprophytes. Here we found that all proteins were present in all six strains; however, there is amino acid sequence variance in LIC12906, LIC13248, LIC10997, LIC12892, LIC12099, LIC12895, which are encoded by genes *LenA, LenF, LenB, Lp29, LipL53 and Lsa27,* respectively. The difference can be natural or due to passage effects (Table A25).

Table A25. Protein homology of adhesion function proteins in six *L. interrogans* strains with reference to strain Fiocruz L1-1

Gene locus (Protein name	Accession	Langkawi	782	1489	898	1548	1530[d]
LIC_10054 (MPL36)	AAS68691.1	100%	99%	100%	100%	100%	100%
LIC_10314 (Lsa63)	AAS68941.1	100%	100%	100%	100%	100%	100%
LIC_10788 (FlaA-1)	AAS69404.1	100%	100%	100%	100%	100%	100%
LIC_10973 (OmpL1)	AAS69584.1	100%	98%	98%	100%	98%	98%
LIC_11194 (citrate lyase)	AAS69801.1	100%	100%	100%	100%	100%	100%
LIC_11436 (Mfn7)	AAS70035.1	100%	99%	100%	100%	99%	100%
LIC_11469 (Lsa20)	AAS70067.1	99%	100%	99%	100%	100%	99%
LIC_11947 (LepA)	AAS70529.1	100%	100%	100%	100%	100%	100%
LIC_11954 (enolase)	AAS70536.1	100%	100%	100%	100%	100%	100%
LIC_11975 (Lsa36)	AAS70551.1	99%	100%	100%	100%	100%	100%
LIC_12238 (rLIC_12238)	AAS70810.1	100%	100%	100%	100%	100%	100%
LIC_12263 (OmpL37)	AAS70835.1	100%	100%	100%	100%	99%	100%
LIC_12407 (Glutamine synthetase)	AAS70976.1	100%	100%	100%	100%	100%	100%
LIC_12730 (rLIC_12730)	AAS71287.1	100%	100%	99%	100%	100%	100%
LIC_12795 (Acetyl-CoA acetyltransferase)	AAS71349.1	100%	100%	100%	100%	100%	100%
LIC_12875 (Ef Tu)	AAS71428.1	100%	100%	100%	100%	100%	100%
LIC_12906 (LfhA/Lsa24/LenA)	AAS71459.1	99%	97%	73%	100%	97%	98%
LIC_12976 (rLIC_12976)	AAS71526.1	100%	100%	100%	100%	100%	100%
LIC_13050 (OmpL47)	AAS71599.1	100%	99%	100%	100%	100%	100%
LIC_13143 (TlyC)	AAS71688.1	100%	100%	100%	100%	100%	100%
LIC_10537 (Mfn9)	AAS69158.1	99%	99%	99%	100%	100%	99%
LIC_12253 (Lsa25)	AAS70825.1	100%	100%	100%	100%	100%	100%
LIC_10714 (Mfn2)	AAS69335.1	100%	99%	100%	100%	100%	100%
LIC_10524 (DnaK)	AAS69145.1	100%	99%	99%	100%	99%	99%
LIC_12315 (LenD)	AAS70886.1	99%	99%	100%	99%	99%	99%
LIC_12690 (Lp95)	AAS71248.1	99%	99%	99%	100%	99%	99%
LIC_13248 (LenF)	AAS71792.1	77%	99%	94%	100%	99%	76%
LIC_13467 (LenE)	AAS72006.1	99%	99%	99%	100%	99%	100%
LIC_13006 (LenC)	AAS71556.1	99%	100%	99%	100%	99%	100%
LIC_10997 (LenB)	AAS69608.1	96%	96%	96%	100%	94%	96%
LIC_10258 (Lsa66)	AAS68886.1	99%	99%	99%	100%	99%	99%
LIC_10731 (Lsa45)	AAS69350.1	100%	100%	100%	100%	100%	100%
LIC_11009 (Lsa26)	AAS69619.1	99%	100%	99%	100%	99%	100%
LIC_11352 (LipL32)	AAS69953.1	100%	100%	100%	100%	100%	100%
LIC_11360 (Lsa23)	AAS69961.1	99%	99%	99%	100%	98%	99%
LIC_11612 (Mfn1)	AAS70207.1	99%	99%	100%	100%	99%	99%
LIC_11834 (Lsa33)	AAS70420.1	100%	100%	100%	100%	100%	100%
LIC_12816 (Conserved hyptothetical protein)	AAS71370.1	100%	100%	100%	100%	100%	100%
LIC_10091 (LipL40)	AAS68725.1	100%	100%	100%	100%	100%	100%
LIC_12892 (Lp29)	AAS71445.1	49%	98%	50%	100%	49%	99%
LIC_10494 (rLIC_10494)	AAS69115.1	100%	46%	46%	100%	99%	99%
LIC_10793 (Lp49)	AAS69409.1	99%	99%	100%	100%	99%	99%
LIC_11087 (Lsa30)	AAS69694.1	100%	100%	100%	100%	99%	100%
LIC_12880 (Lp30)	AAS71433.1	99%	100%	99%	100%	56%	100%
LIC_10645 (Lsa44)	AAS69266.1	99%	99%	99%	100%	100%	99%
LIC_11051 (Mfn6)	AAS69658.1	99%	99%	99%	100%	99%	99%
LIC_12099 (LipL53)	AAS70670.1	92%	91%	98%	100%	91%	90%
LIC_10368 (Lsa21)	AAS68992.1	100%	100%	100%	100%	100%	100%
LIC_12895 (Lsa27)	AAS71448.1	88%	60%	60%	100%	88%	60%

4.5.2.4 Immunodominant proteins

The immunodominant proteins LigA and LigB are two of the most potent extracellular matrix binding proteins. LigA is capable of binding fibronectin, laminin, collagen 1 & IV, while LigB can bind collagen III, elastin, and tropoelastin. LigA and B only present in the low passage and they are surface exposed adhesion molecules (Lin et al., 2010). In this study, we found that protein homology differed in the LigA and LigB proteins as well as in methyltransferase proteins (surface exposed proteins related to the *rfb* locus). The difference in amino acid sequence can be natural or due to passage effects (Table A26).

Table A26. Protein homology of immunodominant proteins in six *L. interrogans* strains with reference to strain Fiocruz L1-130.

Protein	Genebank	Name	Langkawi	782	898	1489	1548	1530d
LigA7-13	NC_005823.1	LigA non-identical region	99%	84%	99%	84%	99%	100%
LigB7-12	NC_005823.1	LigB non-identical region	98%	98%	100%	98%	97%	98%
LIC11352	NC_005823.1	Lipl32	100%	100%	100%	100%	100%	100%
LIC10215	NC_005823.1	TRAM domain protein	100%	100%	100%	100%	100%	100%
LIC11573	NC_005823.1	Putative type II secretion system protein G	100%	100%	100%	100%	100%	100%
LIC11456	NC_005823.1	Putative lipoprotein	100%	100%	100%	100%	100%	100%
LigA/B1-6	NC_005823.1	LigA and LigB Identical region	99%	98%	100%	98%	99%	99%
LIC11335	NC_005823.1	Chaperonin GroEL	100%	100%	100%	100%	100%	100%
LIC11222	NC_005823.1	Tetratricopeptide repeat protein	100%	99%	100%	100%	100%	99%
LIC11389	NC_005823.1	Flagellar protein FlbB	100%	100%	100%	100%	99%	100%
LIC11955	NC_005823.1	Conserve Hypothetical protein	100%	100%	97%	98%	100%	100%
LIC11271	NC_005823.1	Conserve Hypothetical protein	99%	99%	100%	99%	100%	99%
LIC10486	NC_005823.1	Conserve Hypothetical protein	99%	99%	100%	99%	99%	99%
LIC12180	NC_005823.1	Methyltransferase	100%	100%	100%	45%	46%	100%
LIC10191	NC_005823.1	OmpA	100%	100%	100%	100%	100%	100%
LIC20042	NC_005824.1	BatC	99%	99%	100%	100%	99%	100%
LIC20087	NC_005824.1	Membrane protein	100%	100%	100%	100%	99%	99%
LIC12544	NC_005823.1	Transcriptional regulator	100%	100%	100%	100%	100%	100%
LIC11570	NC_005823.1	Putative type II secretion system protein D	100%	100%	100%	100%	100%	100%
LIC20301	NC_005824.1	Hypothetical protein	100%	100%	100%	100%	100%	100%
LIC10524	NC_005823.1	Molecular chaperone DnaK	100%	99%	100%	99%	99%	99%
LIC11437	NC_005823.1	Adenylate/guanylate cyclase	99%	99%	100%	99%	99%	99%
LIC10483	NC_005823.1	Membrane protein PF09851 family	100%	100%	100%	100%	99%	100%
LIC10623	NC_005823.1	Flagellar motor protein MotB	100%	100%	100%	100%	100%	100%

24 proteins were selected due to their high seroreactivity based on study by Lessa-Aquino et al., 2013.

4.5.3 Bacterial Virulence Factors

In *L. interrogans* we detected a total of 76 potential virulence factors displaying sequence homology to those of experimentally proven ones within 19 known bacterial pathogens. Our set of virulence factors consists of 29 defensive, 37 offensive, nine non-specific and one regulation mechanism. The virulence factor most often found in the six strains of *L. interrogans* is AlgI (alginate o-acetyltransferase) of *P.aeruginosa* (PAO1), with a frequency of up to nine copies. This was followed by MtrD (multiple transferable resistance system protein of *Neisseria meningitidis* MC58) with four copies in all six strains. All virulence factors associated with alginate production and regulation are described and discussed separately in the section 'Alginate Biosynthesis'. Thus only 65 virulence factors are described and discussed in this subtopic.

Within the *rfb* locus of the six genomes, several virulence factors from various types of gram-negative and -positive bacteria are located. All four strains (Langkawi, 782, 898 and 1530d) possess similar types of common bacteria virulence factors on their *rfb* loci. These virulent factors are Cap8E, Cap8F, Cap8G, DdhB, DdhA, PseI, PseB, NeuB, NeuC, RffG, BlpC, Per and Fcl. However, strain 1489 and strain 1548 do not show genes related to *Staphylococcus aureus* (Cap8EFG); instead, they have two copies of NeuB related to *Streptoccocus agalactiae* and the PseC protein from *Campylobacter jejuni*. Furthermore, BplC and PseB in these two strains have different product names in their locus tag compared to the other four strains.

Based on the virulence factors that were found in the virulence factors database for bacteria (VFDB), we can infer about the pathogenesis of *L. interrogans.* The virulence factors can be divided basically into four categories, i.e. defensive, offensive, non-specific and regulatory. The defensive mechanism genes can be classified into several subcategories:

i) stress proteins (KatA, ClpC and RecN),
ii) Resistance-Nodulation-Division (RND) efflux system (MtrD)
iii) biofilm formation (AlgI and AlgD)

iv) antiphagocytic capsule (Cap8D, Cap8E, Cap8F, Cap8G, NeuB and NeuC).

For the offensive mechanism, genes that were found are divided into the following subcategories:

i) flagellar motility (FleN, FleQ, FleR/FlrC, FlgG, FlgI, FlhA, FliG and FliM)
ii) endotoxin (BplC, Fcl, Per, RffG, PseI, PseB and PseC)
iii) adherence and chemotaxis (KdsA, KpsF, LpxA, LpxB, LpxC, MbsA, HtpB, RfaD, VhxB, GalE, GmD, BplF, CheA, CheB)
iv) secretion system; CdsN, GspD, GspE, XcpS,
v) toxin (hemolysin); Hlb and HlyB.

For non-specific mechanism, the following are the subcategories:

i) exoenzyme (ColA, SmcL)
ii) mineral uptake (CcmF, MgTB and FeoB)
iii) o-antigen (DdhA, DdhB and Fcl)

In terms of regulatory mechanisms, RelA regulates GTP pyrophosphokinase *relA* (ATP:GTP 3'-pyrophosphotransferase) (PPGPP synthetase I) in *Mycobacterium tuberculosis*.

Protein homology of bacterial virulence factors of the six *L. interrogans* isolates are as shown in Table A27.

Table A27. Protein homology of 65 bacterial virulence factors in six *L. interrogans* strains. Virulence factors with alignment scores of e-value ≤$10e^{-52}$ were taken as significant. Altogether there are 39 offensive virulence factors,16 defensive virulence factors, nine non-specific virulence factors and one regulatory virulence factor.

General Function	Virulence Factor	Genes	Langkawi	782	898	1489	1548	1530d
Offensive VF	lipopolysaccharide biosynthesis protein [LPS] [Bordetella pertussis Tohama I]	bplC	9.00E-72	9.00E-72	8.00E-60	4.00E-58	2.00E-71	1.00E-59
Offensive VF	lipopolysaccharide biosynthesis protein [LPS] [Bordetella pertussis Tohama I]	bplF	1.00E-94	1.00E-94	1.00E-94	1.00E-94	9.00E-95	1.00E-94
Defensive VF	capsular polysaccharide synthesis enzyme Cap8D [Capsule] [Staphylococcus aureus subsp. aureus MW2]	cap8D	e-112	e-112	e-112	e-113	e-112	e-122
Defensive VF	capsular polysaccharide synthesis enzyme Cap8E [Capsule] [Staphylococcus aureus subsp. aureus MW2]	cap8E	e-126	e-126	e-126	9.00E-69	8.00E-69	e-126
Defensive VF	capsular polysaccharide synthesis enzyme Cap8F [Capsule] [Staphylococcus aureus subsp. aureus MW2]	cap8F	8.00E-92	4.00E-92	8.00E-92	*N	*N	6.00E-92
Defensive VF	capsular polysaccharide synthesis enzyme Cap8G [Capsule] [Staphylococcus aureus subsp. aureus MW2]	cap8G	e-115	e-115	e-115	*N	*N	e-115
Non-specific VF	cytochrome c heme lyase subunit CcmF [Cytochrome c muturation (ccm) locus] [Legionella pneumophila subsp.pr	ccmF	3.00E-58	3.00E-58	2.00E-55	3.00E-58	9.00E-64	1.00E-63
Offensive VF	Type III secretion system ATPase [TTSS] [Chlamydia trachomatis D/UW-3/CX]	cdsN	1.00E-96	1.00E-96	1.00E-96	1.00E-96	1.00E-96	1.00E-96
Offensive VF	chemotaxis protein CheA [peritrichous flagella (AI145)] [Yersinia enterocolitica subsp. enterocolitica 8081]	cheA	3.00E-84	3.00E-83	8.00E-84	2.00E-83	8.00E-84	2.00E-78
Offensive VF	chemotaxis histidine kinase [Pse5Ac7Ac, Pse5Ac7Am, Pse8OAc,Pse5Am7AcGlnAc] [Campylobacter jejuni subsp.	cheA	5.00E-93	6.00E-94	6.00E-94	6.00E-94	5.00E-93	5.00E-93
Offensive VF	chemotaxis-specific methylesterase CheB [peritrichous flagella] [Yersinia enterocolitica subsp. Enterocolitica	cheB	1.00E-88	1.00E-88	1.00E-88	1.00E-88	1.00E-88	3.00E-93
Offensive VF	chemotaxis-specific methylesterase [Flagella (VF0430)] [Burkholderia pseudomallei K96243]	cheB	2.00E-81	3.00E-69	3.00E-69	3.00E-69	3.00E-69	3.00E-69
Defensive VF	endopeptidase Clp ATP-binding chain C [ClpC] [Listeria monocytogenes EGD-e]	clpC	0.0	0.0	0.0	0.0	0.0	0.0
Defensive VF	endopeptidase Clp ATP-binding chain C [ClpC] [Listeria monocytogenes EGD-e]	clpC	e-143	0.0	e-142	e-142	e-143	e-143
Defensive VF	endopeptidase Clp ATP-binding chain C [ClpC] [Listeria monocytogenes EGD-e]	clpC	0.0	0.0	0.0	0.0	0.0	0.0
Defensive VF	ATP-dependent Clp protease proteolytic subunit [ClpP] [Listeria monocytogenes EGD-e]	clpP	4.00E-57	4.00E-57	4.00E-57	4.00E-57	4.00E-57	4.00E-57
Non-specific VF	collagenase [kappa-toxin] [Clostridium perfringens str. 13]	colA	2.00E-70	2.00E-70	2.00E-70	2.00E-70	7.00E-71	2.00E-70
Non-specific VF	glucose-1-phosphate cytidylyltransferase [O-antigen][Yersinia enterocolitica subsp. enterocolitica 8081]	ddhA	7.00E-80	7.00E-80	7.00E-80	1.00E-67	1.00E-67	7.00E-80
Non-specific VF	CDP-glucose 4,6-dehydratase [O-antigen] [Yersinia enterocolitica subsp. enterocolitica 8081]	ddhB	8.00E-92	8.00E-92	8.00E-92	4.00E-96	7.00E-96	8.00E-92
Non-specific VF	GDP-fucose synthetase [O-antigen] [Yersinia enterocolitica subsp. enterocolitica 8081]	fcl	e-102	e-102	e-102	7.00E-97	7.00E-97	e-102
Non-specific VF	ferrous iron transporter B [FeoAB] [Legionella pneumophila subsp. pneumophila str. Philadelphia 1]	feoB	2.00E-78	3.00E-78	2.00E-78	3.00E-78	4.00E-78	3.00E-78
Offensive VF	flagellar synthesis regulator FleN [polar flagella][Legionella pneumophila subsp. pneumophila str. Philadelphia 1]	fleN	3.00E-54	3.00E-54	2.00E-54	3.00E-54	3.00E-54	3.00E-54
Offensive VF	transcriptional regulator FleQ [polar flagella] [Legionella pneumophila subsp. pneumophila str. Philadelphia 1]	fleQ	2.00E-68	2.00E-68	2.00E-68	2.00E-68	2.00E-68	2.00E-68
Offensive VF	sigma 54-dependent response regulator [polar flagella] [Legionella pneumophila subsp. pneumophila str.Philadelphi	fleR/flrC	5.00E-83	3.00E-83	5.00E-83	5.00E-83	5.00E-83	5.00E-43
Offensive VF	flagellar basal-body rod protein (flgG) [Pse5Ac7Ac] [Helicobacter pylori 26695]	flgG	1.00E-74	1.00E-74	1.00E-74	1.00E-74	1.00E-74	1.00E-74
Offensive VF	flagellar P-ring protein precursor FlgI [Flagella] [Pseudomonas aeruginosa PAO1]	flgI	6.00E-55	8.00E-55	6.00E-55	6.00E-55	6.00E-55	6.00E-55
Offensive VF	flagellar biosynthesis protein FlhA [Flagella] [Pseudomonas aeruginosa PAO1]	flhA	e-110	e-110	e-110	e-110	e-110	e-110
Offensive VF	flagellar motor switch protein G [Pse5Ac7Ac] [Helicobacter pylori 26695]	fliG	3.00E-60	3.00E-60	3.00E-60	3.00E-60	3.00E-60	3.00E-60
Offensive VF	flagellar motor switch protein FliM [Pse5Ac7Ac] [Helicobacter pylori 26695]	fliM	3.00E-60	3.00E-60	3.00E-60	3.00E-60	3.00E-60	3.00E-60
Offensive VF	UDP-glucose 4-epimerase [LOS] [Haemophilus influenzae Rd KW20]	galE	4.00E-58	1.00E-57	1.00E-57	1.00E-57	1.00E-57	1.00E-57
Offensive VF	GDP-mannose 4,6-dehydratase [LPS] [Brucella melitensis bv. 1 str. 16M]	gmd	e-122	e-122	e-122	e-122	e-122	e-122
Offensive VF	general secretion pathway protein D [T2SS] [Shigella dysenteriae Sd 197]	gspD	2.00E-54	2.00E-54	2.00E-54	2.00E-54	2.00E-54	2.00E-54
Offensive VF	general secretion pathway protein E [T2SS] [Shigella dysenteriae Sd197]	gspE	e-120	e-120	e-120	e-120	e-120	e-120

(cont.)

Offensive VF	beta-hemolysin [-hemolysin] [Staphylococcus aureus subsp. aureus COL]	hlb	2.00E-62	2.00E-57	2.00E-62	2.00E-62	2.00E-57	2.00E-57
Offensive VF	beta-hemolysin [-hemolysin] [Staphylococcus aureus subsp. aureus COL]	hlb	3.00E-55	3.00E-55	3.00E-55	6.00E-52	3.00E-55	3.00E-55
Offensive VF	hemolysin transport protein [Hemolysin] [Escherichia coli O157:H7 str. EDL933]	hlyB	2.00E-56	2.00E-56	8.00E-56	2.00E-56	2.00E-56	2.00E-56
Offensive VF	Hsp60, 60K heat shock protein HtpB [Hsp60] [Legionella pneumophila subsp. pneumophila str. Philadelphia 1]	htpB	1.00E-169	e-169	e-169	e-169	e-169	e-169
Defensive VF	catalase [KatA] [Neisseria meningitidis MC58]	katA	e-157	e-157	e-157	e-157	e-157	e-157
Offensive VF	2-dehydro-3-deoxyphosphooctonate aldolase [LPS] [Brucella melitensis bv.1 str.16M]	kdsA	1.00E-71	1.00E-71	1.00E-71	1.00E-71	1.00E-71	1.00E-71
Offensive VF	D-arabinose 5-phosphate isomerase [Capsule] [Campylobacter jejuni subsp. jejuni NCTC 11168]	kpsF	*5.00E-61*	5.00E-61	5.00E-61	2.00E-61	1.00E-60	2.00E-61
Offensive VF	UDP-N-acetylglucosamine acyltransferase [LOS] [Haemophilus influenzae Rd KW20]	lpxA	2.00E-53	2.00E-53	2.00E-53	2.00E-53	2.00E-53	2.00E-53
Offensive VF	lipid-A-disaccharide synthase [LOS] [Haemophilus influenzae Rd KW20]	lpxB	1.00E-53	1.00E-53	2.00E-53	2.00E-53	1.00E-53	2.00E-53
Offensive VF	UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine deacetylase [LOS] [Haemophilus influenzae Rd KW20]	lpxC	9.00E-53	9.00E-53	9.00E-53	9.00E-53	9.00E-53	9.00E-53
Non-specific VF	Mg2+ transport protein [MgtBC] [Salmonella enterica subsp.enterica serovar Typhimurium str. LT2]	mgtB	e-133	e-147	e-147	e-147	e-146	e-146
Offensive VF	lipid transporter ATP-binding/permease [LOS] [Haemophilus influenzae Rd KW20]	msbA	3.00E-88	2.00E-88	8.00E-88	3.00E-88	3.00E-88	3.00E-88
Offensive VF	lipid transporter ATP-binding/permease [LOS] [Haemophilus influenzae Rd KW20]	msbA	5.00E-84	4.00E-84	4.00E-84	4.00E-84	5.00E-84	4.00E-84
Defensive VF	multiple transferable resistance system protein MtrD [MtrCDE][Neisseria meningitidis MC58]	mtrD	2.00E-56	5.00E-54	2.00E-54	2.00E-56	2.00E-56	2.00E-69
Defensive VF	multiple transferable resistance system protein MtrD [MtrCDE][Neisseria meningitidis MC58]	mtrD	5.00E-56	2.00E-56	e-56	5.00E-54	5.00E-54	1.00E-91
Defensive VF	multiple transferable resistance system protein MtrD [MtrCDE][Neisseria meningitidis MC58]	mtrD	2.00E-69	1.00E-69	5.00E-92	2.00E-69	2.00E-69	5.00E-54
Defensive VF	multiple transferable resistance system protein MtrD [MtrCDE][Neisseria meningitidis MC58]	mtrD	9.00E-92	7.00E-92	5.00E-92	1.00E-91	1.00E-91	2.00E-56
Defensive VF	N-acetyl neuramic acid synthetase NeuB [Capsule][Streptococcus agalactiae 2603V/R]	neuB	5.00E-93	5.00E-93	5.00E-93	7.00E-53	7.00E-53	5.00E-93
Defensive VF	N-acetyl neuramic acid synthetase NeuB [Capsule][Streptococcus agalactiae 2603V/R]	neuB	N	N	N	*2E-91	*2E-91	N
Defensive VF	UDP-N-acetylglucosamine-2-epimerase NeuC [Capsule] [Streptococcus agalactiae 2603V/R]	neuC	5.00E-65	6.00E-64	5.00E-65	*1E-79	*1E-79	5.00E-65
Offensive VF	Perosamine synthetase [LPS] [Brucella melitensis bv. 1 str.16M]	per	8.00E-55	8.00E-55	1.00E-54	6.00E-156	4.00E-160	8.00E-55
Offensive VF	UDP-N-acetylglucosamine 4,6-dehydratase [Pse5Ac7Ac][Helicobacter pylori 26695]	pseB	8.00E-66	8.00E-66	7.00E-66	*2E-97	*2E-97	8.00E-66
Offensive VF	C4 aminotransferase [Pse5Ac7Ac, Pse5Ac7Am, Pse8OAc, Pse5Am7AcGlnAc] [Campylobacter jejuni subsp. jejuni	pseC	N	N	N	*4E-65	*4E-65	N
Offensive VF	N-acetylneuraminic acid synthetase [Pse5Ac7Ac, Pse5Ac7Am, Pse8OAc, Pse5Am7AcGlnAc] [Campylobacter jejur	psel	8.00E-52	6.00E-52	8.00E-52	*8E-65	*7E-65	8.00E-52
Defensive VF	DNA repair protein RecN [RecN] [Neisseria meningitidis MC58]	recN	7.00E-66	7.00E-66	7.00E-66	7.00E-66	7.00E-66	7.00E-66
Regulation	probable GTP pyrophosphokinase RelA (ATP:GTP 3'-pyrophosphotransferase) (PPGPP synthetase I) ((P)PPGPP s	relA	e-149	e-149	e-149	e-149	e-149	e-149
Offensive VF	*ADP-L-glycero-D-mannoheptose-6-epimerase [LOS] [Haemophilus influenzae Rd KW20*	rfaD	2.00E-57	4.00E-58	4.00E-58	4.00E-58	3.00E-57	1.00E-57
Offensive VF	dTDP-glucose 46-dehydratase [LOS] [Haemophilus influenzae Rd KW20]	rffG	2.00E-97	2.00E-97	2.00E-97	2.00E-97	1.00E-97	2.00E-97
Non-specific VF	sphingomyelinase-c [SMase] [Listeria ivanovii str. ATCC 19119]	smcL	1.00E-59	1.00E-59	2.00E-59	1.00E-59	1.00E-59	1.00E-59
Non-specific VF	sphingomyelinase-c [SMase] [Listeria ivanovii str. ATCC 19119]	smcL	1.00E-69	1.00E-69	4.00E-69	1.00E-69	1.00E-69	1.00E-69
Offensive VF	general secretion pathway protein F [xcp secretion system] [Pseudomonas aeruginosa PAO1]	xcpS	2.00E-56	2.00E-56	2.00E-56	2.00E-56	2.00E-56	2.00E-56
Offensive VF	phosphomannomutase [LOS] [Haemophilus influenzae Rd KW20]	yhxB/manB	1.00E-62	6.00E-63	6.00E-63	2.00E-62	6.00E-63	2.00E-65

N= not signsificant i.e. e-value of alignment result is larger than e-^{52} or alignment score lesser than 200
*Similarities found only between strain1489 and strain 1548.
Green, defensive virulence factors; light red, offensive virulence factors; yellow, non-specific virulence factors

4.5.4 Alginate Biosynthesis

Alginate biosynthesis is a survival mechanism for leptospires in the environment. This system is absent in the pathogenic species *L. borgpetersenii,* which explains the inability of this species to survive in the environment. The saprophytic leptospires generate biofilm much faster compared to pathogenic leptospires, whereas the low passage leptospires have more efficient biofilm synthesis compared to high-passage leptospires (Ristow et al., 2008). In this study, the alginate biosynthesis and transport proteins of six *L. interrogans* genomes were queried for protein homology against those of alginate biosynthesis of *P.aeruginosa* (Table A28).

Table A28. Protein homology of in alginate structural biosynthesis, regulatory and genotypic switching in six *L. interrogans* strains with reference to *P. aeruginosa* PA01. 11 proteins associated with structural biosynthetic, three proteins from regulatory proteins and three proteins from genotypic switching from *P. aeruginaosa* PA01 were not found in the six strains of *L. interrogans.*

Gene cluster	NCBI	Gene	Gene product	Langkawi	782	898	1489	1548	1530d
Structural/ biosynthetic	ARG85621.1	algD	GDP-mannose dehydrogenase	6,00E-63	5,00E-54	3,00E-63	5,00E-54	5,00E-54	5,00E-54
	OGX66506.1	alg8	subunit of alginate polymerase	X	X	X	X	X	X
	PTC38438.1	alg44	subunit of alginate polymerase	3	X	X	X	X	X
	AVV63678.1	algK	subunit of protein scafold	X	X	X	X	X	X
	PTC38436.1	algE(alg76)	porin-like OM protein	X	X	X	X	X	X
	ARG85616.1	algG	mannuronan C-5 epimerase	X	X	X	X	X	X
	AAA91126.1	algX(alg60)	subunit of protein scafold	3.6	4	3.4	3.4	3.6	3.6
	ARG85614.1	algL	alginate lyase	X	X	X	X	X	X
	ARG85613.1	algI	acetylase	1,00E-90	1,00E-90	1,00E-90	1,00E-90	1,00E-90	1,00E-90
		algI	acetylase	2,00E-83	9,00E-84	2,00E-83	2,00E-83	2,00E-83	2,00E-83
		algI	acetylase	3,00E-83	2,00E-83	3,00E-83	4,00E-83	5,00E-83	3,00E-83
		algI	acetylase	2,00E-80	2,00E-80	2,00E-80	2,00E-80	2,00E-80	2,00E-80
		algI	acetylase	5,00E-80	4,00E-80	2,00E-80	4,00E-80	5,00E-80	4,00E-80
		algI	acetylase	6,00E-76	4,00E-77	4,00E-77	1,00E-76	3,00E-76	4,00E-76
		algI	acetylase	2,00E-73	1,00E-73	1,00E-73	1,00E-73	9,00E-74	1,00E-73
		algI	acetylase	4,00E-73	4,00E-73	4,00E-73	4,00E-73	3,00E-73	4,00E-73
		algI	acetylase	2,00E-68	2,00E-68	3,00E-64	3,00E-64	2,00E-68	3,00E-64
		algI	acetylase	7,00E-64	8,00E-64	2,00E-63	2,00E-63	9,00E-64	7,00E-64
		algI	acetylase	X	X	X	X	X	X
		algI	acetylase	X	X	X	X	X	X
	BAT64151.1	algJ	acetylase	X	X	X	X	X	X
	BAT64150.1	algF	acetylase	X	X	X	X	X	X
	BAR66409.1	algA	PMI-GMP	2,00E-60	1,00E-61	3,00E-62	2,00E-61	3,00E-61	3,00E-61
	ARG90425.1	algC	PMM	5,00E-33	3,00E-33	3,00E-33	7,00E-33	3,00E-33	3,00E-33
Regulatory	AAC44751.1	algZ(fimS)	RHH DNA-binding protein	X	X	X	2.0	1.8	X
	PTC33675.1	algR(algR1)	respnse regulator protein of TCSTS	8,00E-12	9,00E-12	6,00E-12	2,00E-12	1.07	2,00E-12
	BAR70739.1	algQ (algR2)	cognate sensor Kinase	X	X	X	X	X	X
	SIP55675.1	algP(algR3)	histone-like protein	X	X	X	X	X	X
	PTC33420.1	algB	NtrC subclass of TCSTS	9,00E-90	1,00E-89	8e_90	1,00E-89	9,00E-90	9,00E-90
Genotypic switching	BAT67364.1	algT(algU)	sigma factor 22	8,00E-30	7,00E-30	8,00E-30	7,00E-30	6,00E-30	6,00E-30
	ABQ41381.1	mucA(algS)	anti-22 factor	X	X	X	X	X	X
	AAA87633.1	mucB(algN)	anti-22 factor	X	3.6	X	3.9	X	X
	ABV00671.1	mucC(algM)	homologue of PhoORF4 product	X	X	X	X	X	X
	AAK11276.1	mucD(algW/Y)	homologue of serine protease (HtrA)	2,00E-70	2,00E-71	2,00E-71	2,00E-71	2,00E-71	1,00E-72

X, amino acid sequence homology was not found

4.6 Comparative Transcriptomic Analysis

A transcriptomic comparison was performed between two *L. interrogans* strains of extreme virulence, i.e. strain Langkawi being the most virulence and strain 1530d being the least virulence, in proteins associated with motility, chemotaxis and adhesion and immunodominance in reference to *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 as in Fouts et al., 2016. Furthermore, a transcriptomic comparison to 65 bacterial virulence factors from VFDB and alginates biosynthesis in reference to *P.aeruginosa* strain PA01 was also performed.

4.6.1 Motility

36 out of 51 (70.5%) proteins associated with motility were transcribed differentially between the two strains. 30 out of 36 (83.3%) proteins of Langkawi were expressed at log2 fold change (FC) less than -1, in strain Langkawi compared to that of strain 1530d (highlighted in light red). 24 out of 30 (80%) of these differential expressions were significant with p-value <0.05. As *flgG, flgA, flgH and flgI* are clustered genes and transcribed in a similar manner, most likely they are encoded by a polycistronic message. Six out of 36 (16.7%) genes encoding proteins, i.e. flagellar hook-basal body complex protein, flagellar protein, flagellar assembly protein / ATPase complex, *FlhB*-like protein Type III secretion, flagellar filament core protein *FlaB*, flagellar protein (modular protein), were expressed higher than two-fold in strain Langkawi compared to strain 1530d (highlighted in green) (Table A29).

Table A29. Transcriptomic comparison of motility genes encoding structural proteins between three biological replicates of strain Langkawi and strain 1530d based on TPM value. Generally, the expressions of 30 genes were higher in strain 1530d than in strain Langkawi. 24 out of 30 (80%) of the differential expressions were significant p-value <0.05.

LIC ORFs	Gene	Accession No.	Annotation	Langkawi_1	Langkawi_2	Langkawi_3	1530d_1	1530d_2	1530d_3	Average_ Langkawi	Average_ 1530d	Fold Change (FC)	Log2 FC	T-test
BASAL BODY														
LIC10297	*fliE*	*AAS68924.1*	Flagellar hook-basal body complex protein	133	118	135	42	61	62	129	55	2,33182	1,22146	0,001151
LIC10298	*flgC*	*AAS68925.1*	Flagellar basal-body rod protein	27	30	35	49	93	76	30	73	0,41902	-1,25491	0,074309
LIC10299	*flgB*	*AAS68926.1*	Flagellar basal-body rod protein	26	31	33	47	68	70	30	62	0,48579	-1,04161	0,044985
LIC10621	*flbD*	*AAS69242.1*	Flagellar protein	3	1	1	16	17	15	2	16	0,09977	-3,32522	0,000118
LIC10622	*motA*	*AAS69243.1*	Chemotaxis protein	3	2	3	18	28	28	3	25	0,11545	-3,11463	0,020293
LIC12931	*motA1*	*AAS71482.1*	Chemotaxis protein	10	9	11	19	24	27	10	23	0,42726	-1,22680	0,024909
LIC10623	*motB*	*AAS69244.1*	Chemotaxis protein	5	7	6	22	31	31	6	28	0,21584	-2,21196	0,017691
LIC10624	*fliL*	*AAS69245.1*	Flagellar protein	7	8	9	22	44	37	8	34	0,23004	-2,12002	0,055115
LIC11187	*flgD*	*AAS69794.1*	Flagellar hook assembly scaffolding protein	18	20	23	121	166	180	20	156	0,12984	-2,94518	0,016131
LIC11324	*flgG*	*AAS69927.1*	Flagellar basal-body rod protein	16	11	11	21	34	40	13	32	0,39953	-1,32363	0,065160
LIC11325	*flgA*	*AAS69928.1*	Chaperone P-ring formation protein	48	24	28	49	105	115	33	90	0,37126	-1,42951	0,097786
LIC11326	*flgH*	*AAS69929.1*	Flagellar L-ring protein	5	5	6	24	39	42	5	35	0,14929	-2,74383	0,031425
LIC11327	*flgI*	*AAS69930.1*	Flagellar P-ring protein - hook-associated	5	5	5	21	34	40	5	32	0,16056	-2,63884	0,039887
LIC11389	*flbB*	*AAS69989.1*	Flagellar protein	58	67	77	22	32	33	67	29	2,33649	1,22434	0,007890
LIC11390	*fliJ*	*AAS69013.1*	Chaperone-like protein export apparatus	2	1	0	9	4	4	1	5	0,18080	-2,46756	0,113623
LIC11391	*fliI*	*AAS69014.1*	Flagellum-specific ATP synthase/ATPase	1	1	1	5	6	6	1	6	0,17967	-2,47657	0,012109
LIC11394	*fliH*	*AAS69017.1*	Flagellar assembly protein/ATPase complex	6	6	9	3	3	1	7	2	2,87656	1,52435	0,018606
LIC11393	*fliG1*	*AAS69993.1*	Flagellar motor switch protein/C ring	8	11	9	20	35	32	9	29	0,32538	-1,61981	0,046756
LIC11071	*flgM*	*AAS69678.1*	Anti sigma factor for FliA	198	303	335	421	806	856	279	695	0,40122	-1,31755	0,083500
LIC13452	*flgN*	*AAS71992.1*	Chaperone export apparatus	11	13	11	103	145	173	12	140	0,08455	-3,56405	0,024032
LIC13449	*fliW*	*AAS71989.1*	Flagellar assembly factor	11	14	13	49	62	77	12	62	0,19742	-2,34067	0,024372
LIC11562	*flhX*	*AAS70158.1*	FlhB-like protein Type III secretion	133	118	135	62	53	59	129	58	2,21040	1,14431	0,001514
HOOK														
LIC13450	*flgL*	*AAS71990.1*	Flagellar hook-associated protein	11	12	12	62	79	93	12	78	0,15350	-2,70373	0,017655
LIC13451	*flgK*	*AAS71991.1*	Flagellar hook-associated protein	9	11	12	65	91	105	11	87	0,12363	-3,01593	0,022173
LIC12035	*flhO*	*AAS70610.1*	Hook protein/FliG	13	10	11	110	107	123	12	113	0,10171	-3,29753	0,001950
LIC11188	*flgE*	*AAS69795.1*	Flagellar hook protein	20	23	23	105	156	170	22	144	0,15296	-2,70873	0,024923
LIC10723	*fliD*	*AAS69344.1*	Filament cap	44	47	50	192	229	266	47	229	0,20544	-2,28320	0,013259
LIC11186	*fliK*	*AAS69793.1*	Flagellar hook-length control protein	17	17	18	72	85	95	17	84	0,20613	-2,27835	0,009499
LIC11328	*flgJ*	*AAS69931.1*	Peptidoglycan hydrolase/rod cap	8	8	6	21	24	24	7	23	0,31936	-1,64676	0,000640
FILAMENT														
LIC10788	*flaA*	*AAS69404.1*	Flagellar filament outer layer protein A	8	8	6	361	483	525	7	456	0,01631	-5,93784	0,011723
LIC10787	*flaA2*	*AAS69403.1*	Flagellar filament outer layer protein A	67	53	59	215	220	226	60	221	0,27138	-1,88159	0,000013
LIC11531	*flaB*	*AAS70127.1*	Flagellar filament core protein FlaB	32	26	29	164	223	190	29	192	0,15204	-2,71746	0,010007
LIC11890	*flaB2*	*AAS70476.1"*	Flagellar filament core protein FlaB	137	180	198	1795	3134	3563	172	2831	0,06059	-4,04480	0,037682
LIC11889	*flaB3*	*AAS70475.1*	Flagellar filament core protein FlaB	155	193	206	2192	3324	3605	185	3040	0,06083	-4,03900	0,022008
LIC11532	*flaB4*	*AAS70128.1*	Flagellar filament core protein FlaB	827	927	982	141	216	214	912	190	4,79459	2,26141	0,000666
LIC11657	*fliS*	*AAS70250.1*	Flagellar protein (modular protein)	214	214	222	99	72	84	217	85	2,55542	1,35356	0,001320

Green, indicates upregulation ; light red, indicates downregulation; yellow, p-value <0.05

4.6.2 Chemotaxis

18 out of 24 (75%) genes associated with chemotaxis were transcribed differentially between the two strains. 15 out of 18 (83.3%) genes of Langkawi were expressed at log_2FC of more than 1, in strain Langkawi compared to that of strain 1530d. 15 out of 15 (100%) of these differential expressions were significant with p-value <0.05. As *cheB3, cheD1, cheW, mcp, cheA2* are clustered genes and transcribed in a similar manner, most likely they are encoded by a polycistronic message. Genes that encoded methyl-accepting chemotaxis protein (LIC12482), anti-sigma factor antagonist (LIC12482) and methyl-accepting chemotaxis I (LIC10685) were transcribed at log_2FC lower than -1, in strain Langkawi as compared to strain 1530d (Table A30).

Table A30. Transcriptomic comparison of genes encoding chemotaxis genes between three biological replicates of strain Langkawi and strain 1530d based on TPM value. Generally, 15 genes were more highly expressed in strain Langkawi. 16 out of 18 (88.8%) of the differential expressions were significant at a p-value <0.05.

LIC ORFs	Gene	Accession No.	Annotation	Langkawi_1	Langkawi_2	Langkawi_3	1530d_1	1530d_2	1530d_3	Average_ Langkawi	Average_ 1530d	Fold Change (FC)	Log2 FC	T-test
LIC10685	*mcp*	*AAS69306.1*	Methyl-accepting chemotaxis protein	3	4	4	10	20	19	4	17	0,22124	-2,1763	0,05189
LIC11216	*mcp*	*AAS69822.1*	Methyl-accepting chemotaxis protein	266	235	249	21	24	29	250	24	10,24651	3,3571	0,00086
LIC11520	*cheB3*	*AAS70116.1*	Chemotaxis response regulator protein	15	17	16	2	4	5	16	4	4,11283	2,0401	0,00041
LIC11521	*cheD1*	*AAS70117.1*	Chemoreceptor glutamine deamidase	42	43	47	5	9	11	44	9	5,05355	2,3373	0,00014
LIC11522	*cheW*	*AAS70118.1*	Chemotaxis protein	42	49	50	3	6	5	47	5	9,86370	3,3021	0,00213
LIC11523	*mcp*	*AAS70119.1*	Chemotaxis protein	75	97	97	4	7	6	90	6	15,63632	3,9668	0,00662
LIC11524	*cheA2*	*AAS70120.1*	Chemotaxis protein histidine kinase	28	34	35	3	5	4	32	4	8,38184	3,0673	0,00347
LIC11525	-	AAS70121.1	Anti sigma factor antagonist	34	43	42	77	121	117	40	105	0,37875	-1,4007	0,03736
LIC11526	*cheY*	*AAS70122.1*	Chemotaxis response regulator	50	60	55	4	9	8	55	7	7,70105	2,9451	0,00107
LIC11871	*cheR*	*AAS70457.1*	Chemotaxis protein methyltransferase	67	59	73	15	19	21	66	18	3,63435	1,8617	0,00228
LIC12061	*cheB2*	*AAS70632.1*	Chemotaxis protein-glutamate methylesterase	21	16	23	6	10	8	20	8	2,52343	1,3354	0,01307
LIC12062	*cheR1*	*AAS70633.1*	Chemotaxis protein	48	56	61	8	12	13	55	11	4,91117	2,2961	0,00240
LIC12240	*cheW2*	*AAS70812.1*	Chemotaxis protein	56	64	67	3	13	13	62	10	6,41559	2,6816	0,00034
LIC12482	*mcp*	*AAS71047.1*	Methyl-acceptingchemotaxis I	4	4	4	6	11	10	4	9	0,46261	-1,1121	0,07876
LIC12500	*mcpB*	*AAS71065.1*	Chemoreceptor transmembrane protein	38	47	52	13	19	21	46	18	2,59802	1,3774	0,00653
LIC12921	*mcpA*	*AAS71473.1*	Methyl-acceptingchemotaxis protein	60	39	44	19	22	18	48	20	2,45175	1,2938	0,04383
LIC13183	*mcp*	*AAS71728.1*	Methyl-accepting chemotaxis-like	2673	2851	2839	231	360	333	2788	308	9,04556	3,1772	0,00001
LIC13401	*mcp*	*AAS71941.1*	Methyl-accepting chemotaxis protein	38	47	52	13	19	21	46	18	2,59802	1,3774	0,00653

Green, indicates upregulation ; light red, indicates downregulation, yellow, indicates p-value <0.05.

4.6.3 Adhesion

18 out of 49 (36.7%) genes encoding adhesion proteins were transcribed differentially between the two strains. 14 out of 18 (77.7%) proteins of Langkawi were expressed at $\log_2$FC of lower than -1, in strain Langkawi compared to strain 1530d. Only four genes encoding the extracellular matrix ligand proteins, i.e. LIC13467 (*lenE*), LIC11469 (*lsa20*), LIC11834 (*lsa33*) and LIC12892 (*lp29*) were transcribed at $\log_2$FC higher than 1 in strain Langkawi as compared to strain 1530d, and three out of four (75%) of these expressions were significant at a p-value < 0.05 (Table A31).

Table A31 Transcriptomic comparison of genes encoding adhesion proteins between three biological replicates of strain Langkawi and strain 1530d based on TPM value. Adhesion genes included in the analyses encode for ligands to the extracellular matrix, plasminogens, factor H, factor H-like and C4 binding protein. 14 genes were expressed more highly in strain 1530d than in strain Langkawi. 12 out of 18 (66.6%) of these differential expressions were significant at a p-value < 0.05.

Gene locus (Protein name)	Accession	Ligands*	Langkawi_1	Langkawi_2	Langkawi_3	1530d_1	1530d_2	1530d_3	Average_ Langkawi	Average_ 1530d	Fold Change (FC)	Log2FC	T-test
LIC_10054 (MPL36)	AAS68691.1	PLG	22	16	17	66	72	68	18	69	0,26509	-1,9154	3,770E-05
LIC_10788 (FlaA-1)	AAS69404.1	PLG	58	57	70	361	483	525	62	456	0,13487	-2,8904	1,462E-02
LIC_11194 (citrate lyase)	AAS69801.1	PLG	23	27	31	56	101	85	27	81	0,33625	-1,5724	5,288E-02
LIC_11469 (Lsa20)	AAS70067.1	ECM, PLG	4	7	6	5	1	0	6	2	2,84605	1,5090	1,017E-01
LIC_12263 (OmpL37)	AAS70835.1	ECM	7	6	7	21	31	30	7	27	0,24326	-2,0395	2,467E-02
LIC_12407 (Glutamine synthetase)	AAS70976.1	PLG	10	9	12	39	88	75	10	67	0,15612	-2,6792	6,153E-02
LIC_12730 (rLIC_12730)	AAS71287.1	PLG	9	9	9	17	27	27	9	24	0,38086	-1,3927	4,639E-02
LIC_12875 (Ef Tu)	AAS71428.1	ECM, PLG, FH	91	120	141	584	1159	1091	118	945	0,12446	-3,0063	4,400E-02
LIC_10537 (Mfn9)	AAS69158.1	ECM	2	2	2	10	15	14	2	13	0,13244	-2,9166	1,994E-02
LIC_10714 (Mfn2)	AAS69335.1	ECM	4	5	4	10	14	15	4	13	0,32364	-1,6275	3,020E-02
LIC_13467 (LenE)	AAS72006.1	ECM	344	367	357	8	10	9	356	9	40,03381	5,3231	3,060E-04
LIC_13006 (LenC)	AAS71556.1	ECM	7	8	7	23	36	29	7	29	0,24249	-2,0440	2,716E-02
LIC_11834 (Lsa33)	AAS70420.1	ECM, PLG, C4BP	109	123	117	30	48	41	117	40	2,94649	1,5590	4,259E-04
LIC_12892 (Lp29)	AAS71445.1	PLG	9	9	10	3	5	5	9	4	2,20992	1,1440	7,639E-03
LIC_10793 (Lp49)	AAS69409.1	PLG	4	4	4	9	13	12	4	11	0,33558	-1,5753	2,304E-02
LIC_11087 (Lsa30)	AAS69694.1	ECM, PLG, C4BP	1	1	1	2	2	2	1	2	0,35639	-1,4885	3,103E-02
LIC_11051 (Mfn6)	AAS69658.1	ECM	3	4	5	6	13	12	4	10	0,41459	-1,2703	1,146E-01
LIC_12895 (Lsa27)	AAS71448.1	ECM	5	7	7	67	107	104	7	93	0,07077	-3,8207	2,106E-02

*ECM = extracellular matrix; PLG = plasminogen; FH = factor H; FLH-1 = factor H-like; C4BP = C4 binding protein

Green, indicates upregulation ; light red, indicates downregulation; yellow, indicates p-value < 0.05.

4.6.4 Immunodominance

10 out of 24 (41.6%) genes encoding immunodominance proteins were transcribed differently between the two strains. Seven out of ten (70%) genes were expressed at log_2FC lower than -1 in strain Langkawi than in strain 1530d. Four out of ten (40%) of these differential expressions were significant at a p-value <0.05. Three genes encoding membrane protein PF09851 family protein (LIC10483), tetratricopeptide repeat protein (LIC11222) and flagellar protein FlbB (LIC11389) respectively were were expressed at log_2FC higher than 1 in strain Langkawi than in strain 1530d, and all expressions were significant at a p-value <0.05 (Table A32).

Table A32. Transcriptomic comparison of genes encoding immunodominant proteins between three biological replicates of each strain Langkawi and strain 1530d based on TPM value. The genes encoding seven proteins were expressed more highly in strain 1530d, while three other genes were expressed higher in strain Langkawi. Seven out of ten (70%) of these differential expressions were significant at a p-value <0.05.

LIC ORFs	Annotation	Langkawi_1	Langkawi_2	Langkawi_3	1530d_1	1530d_2	1530d_3	Average_ Langkawi	Average_ 1530d	Fold Change (FC)	Log2FC	T-test
LigA7-13	LigA non-identical region	5	5	7	9	13	14	6	12	0,48023	-1,05820	0,05661
LIC11456	Putative lipoprotein	31	17	21	83	79	72	23	78	0,29456	-1,76335	0,00084
LigA/B1-6	LigA and LigB Identical region	5	5	7	9	13	14	6	12	0,48023	-1,05820	0,05661
LIC11335	Chaperonin GroEL	32	37	38	168	278	275	36	240	0,14842	-2,75226	0,02931
LIC11222	Tetratricopeptide repeat protein	25	17	19	8	9	7	20	8	2,54520	1,34778	0,02421
LIC11389	Flagellar protein FlbB	58	67	77	22	32	33	67	29	2,33649	1,22434	0,00789
LIC11271	Conserve Hypothetical protein	14	18	18	24	40	38	17	34	0,49485	-1,01495	0,06950
LIC12544	Transcriptional regulator	23	24	29	47	78	75	25	67	0,37610	-1,41082	0,04638
LIC10483	Membrane protein PF09851 family	139	184	201	33	49	52	175	45	3,91206	1,96793	0,01361
LIC10623	Flagellar motor protein MotB	5	7	6	22	31	31	6	28	0,21584	-2,21196	0,01769

Green, indicates upregulation; light red, indicates downregulation; yellow, indicates p-value < 0.05.

4.6.5 Bacterial Virulence Factors

25 out of 65 (38.4%) genes encoding 65 bacterial virulence factors were transcribed differentially between the two strains. 13 out of 25 (52%) mRNAs of Langkawi were expressed at log_2FC higher than 1 in strain Langkawi than in strain 1530d, and all expressions are significant at p-value <0.05. These genes are; *cheA* (encode chemotaxis protein), *cheB* (encode chemotaxis-specific methylesterase), two copies of *clpC* (encode proteins of endopeptidase ATP-binding chain C), *clpP* (encode ATP-dependent protease proteolytic), *hlb* (encode beta-hemolysin), *hlyB* (encode hemolysin transport protein), *katA* (encode catalase) mtrD(encode multiple transferable resistance system protein), *smcL* (encode sphyngomyelinase), *mgtB*,(encode Mg2+ transport protein), *lpxC* (encode DP-3-OR-3 hydroxymyristoyl-N-acetylglucosamine deacrtylase). Twelve out of 25 (48%) of the genes were expressed at log_2FC lower than -1, and seven out of twelve (58.3%) of these expressions are significant at p-value <0.05. These genes are *bplC, bplF, flgG, flgI, galE, gmd, htpB, lpxA, msbA, mtrD, recN* and *rffG* (Table A33).

Table A33. Transcriptomic comparison of genes encoding bacterial virulence proteins between three biological replicates each of strain Langkawi and strain 1530d based on TPM value. Transcriptomic analyses were performed on mRNA isolated from strain Langkawi and strain 1530d based on locus tags which encoded 65 bacterial virulence proteins of significant protein homology. 13 of the virulence factor genes were expressed two-fold higher in Langkawi (green) than in 1530d (light red). 20 out of 25 (80%) of these differential expressions are significant at p-value <0.05.

Virulence Factor	Genes	Langkawi_1	Langkawi_2	Langkawi_3	1530d_1	1530d_2	1530d_3	Average_ Langkawi	Average_ 1530d	Fold Change (FC)	Log2 FC	T-test
lipopolysaccharide biosynthesis protein [LPS] [Bordetella pertussis Tohama I]	bplC	7	6	6	32	57	51	7	47	0,1407097	-2,829206	0,032640
lipopolysaccharide biosynthesis protein [LPS] [Bordetella pertussis Tohama I]	bplF	4	5	6	29	46	53	5	43	0,1201672	-3,056885	0,035566
chemotaxis protein CheA [peritrichous flagella (AI145)] [Yersinia enterocolitica subsp. enterocolitica 8081]	cheA	28	34	35	3	5	4	32	4	8,3818401	3,067267	0,003468
chemotaxis-specific methylesterase CheB [peritrichous flagella] [Yersinia enterocolitica subsp. Enterocolitica	cheB	15	17	16	2	4	5	16	4	4,1128281	2,040131	0,000410
endopeptidase Clp ATP-binding chain C [ClpC] [Listeria monocytogenes EGD-e]	clpC	729	860	969	97	156	157	853	137	6,2342021	2,640205	0,005959
endopeptidase Clp ATP-binding chain C [ClpC] [Listeria monocytogenes EGD-e]	clpC	70	98	97	28	44	47	88	40	2,2154943	1,147629	0,015146
ATP-dependent Clp protease proteolytic subunit [ClpP] [Listeria monocytogenes EGD-e]	clpP	206	170	189	66	89	89	188	82	2,3040431	1,204168	0,001724
flagellar basal-body rod protein (flgG) [Pse5Ac7Ac] [Helicobacter pylori 26695]	flgG	16	11	11	21	34	40	13	32	0,3995279	-1,323632	0,065160
flagellar P-ring protein precursor FlgI [Flagella] [Pseudomonas aeruginosa PAO1]	flgI	5	5	5	21	34	40	5	32	0,1605574	-2,638839	0,039887
UDP-glucose 4-epimerase [LOS] [Haemophilus influenzae Rd KW20]	galE	6	9	7	22	46	42	8	36	0,2063534	-2,276811	0,055212
GDP-mannose 4,6-dehydratase [LPS] [Brucella melitensis bv. 1 str. 16M]	gmd	7	8	7	22	38	35	7	32	0,2235411	-2,161388	0,038157
beta-hemolysin [-hemolysin] [Staphylococcus aureus subsp. aureus COL]	hlb	16	17	17	5	8	6	17	6	2,6894803	1,427327	0,001251
hemolysin transport protein [Hemolysin] [Escherichia coli O157:H7 str. EDL933]	hlyB	105	117	124	11	17	17	116	15	7,7714833	2,958190	0,001125
Hsp60, 60K heat shock protein HtpB [Hsp60] [Legionella pneumophila subsp. pneumophila str. Philadelphia 1]	htpB	32	37	38	168	278	275	36	240	0,1484186	-2,752256	0,029314
catalase [KatA] [Neisseria meningitidis MC58]	katA	545	629	659	85	121	132	611	113	5,4277831	2,440363	0,001598
UDP-N-acetylglucosamine acyltransferase [LOS] [Haemophilus influenzae Rd KW20]	lpxA	9	12	12	22	44	41	11	36	0,3091802	-1,693480	0,065539
UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine deacetylase [LOS] [Haemophilus influenzae Rd KW20]	lpxC	190	227	226	81	118	119	214	106	2,0262459	1,018809	0,003457
Mg2+ transport protein [MgtBC] [Salmonella enterica subsp.enterica serovar Typhimurium str. LT2]	mgtB	21	26	27	9	13	14	25	12	2,0826795	1,058441	0,006676
lipid transporter ATP-binding/permease [LOS] [Haemophilus influenzae Rd KW20]	msbA	4	4	5	9	15	16	5	13	0,3409313	-1,552447	0,039930
multiple transferable resistance system protein MtrD [MtrCDE][Neisseria meningitidis MC58]	mtrD	6	7	8	12	19	19	7	17	0,4136884	-1,273384	0,054444
multiple transferable resistance system protein MtrD [MtrCDE][Neisseria meningitidis MC58]	mtrD	173	217	219	35	59	59	203	51	3,9815116	1,993316	0,002942
multiple transferable resistance system protein MtrD [MtrCDE][Neisseria meningitidis MC58]	mtrD	15	15	15	5	7	7	15	6	2,3341231	1,222881	0,009385
DNA repair protein RecN [RecN] [Neisseria meningitidis MC58]	recN	21	19	20	264	276	331	20	290	0,0702638	-3,831075	0,005719
dTDP-glucose 46-dehydratase [LOS] [Haemophilus influenzae Rd KW20]	rffG	6	7	7	9	18	16	7	14	0,4818041	-1,053481	0,102694
sphingomyelinase-c [SMase] [Listeria ivanovii str. ATCC 19119]	smcL	5	5	5	1	2	1	5	2	3,1684031	1,663756	0,000280

Green, indicates upregulation ; light red, indicates downregulation; yellow, indicates p-value <0.05.

4.6.6 Alginate biosynthesis

13 out of 18 (72.2%) genes associated with alginate biosynthesis in reference to *P. aeruginosa* strain PA01 were transcribed differentially between the two strains. 10 out of 13 (76.9%) genes were expressed at $\log_2$FC lower than -1 in strain Langkawi than in strain 1530d. Seven out of ten (70%) of these differential expressions were significant at a p-value <0.05. Only three out of 13 (23%) genes were expressed at $\log_2$FC of at least 1 higher in strain Langkawi compared to strain 1530d, and all expressions were significant at a p-value <0.05. These three *algI* genes copies encoded acetylase (*algI*) (Table A33).

Table A34. Transcriptomic comparison of genes involved in alginate biosynthesis between three biological replicates of each strain Langkawi and strain 1530d based on TPM value. The alginates proteins locus tag was defined with reference to that of *P. aeruginosa* strain PA01. Generally, these genes were transcribed more highly in strain 1530d than in strain Langkawi. Ten out of thirteen (76.9%) of these expressions were significant at a p-value <0.05.

Gene cluster	NCBI	Gene	Gene product	Langkawi_1	Langkawi_2	Langkawi_3	1530d_1	1530d_2	1530d_3	Average_ Langkawi	Average_ 1530d	Fold Change (FC)	Log2 FC	T-test
Structural/ biosynthetic	ARG85621.1	algD	GDP-mannose dehydrogenase	5	6	6	31	46	49	5	42	0,12800	-2,9658	0,0223
	AAA91126.1	algX(alg60)	subunit of protein scafold	11	23	26	36	81	82	20	66	0,29735	-1,7498	0,0820
	ARG85613.1	algI	acetylase	5	6	6	31	46	49	5	42	0,12800	-2,9658	0,0223
	ARG85613.1	algI	acetylase	57	82	79	5	7	6	73	6	12,38482	3,6305	0,0124
	ARG85613.1	algI	acetylase	68	117	131	6	10	10	105	9	12,01508	3,5868	0,0359
	ARG85613.1	algI	acetylase	63	78	72	6	13	13	71	11	6,63989	2,7312	0,0010
	ARG85613.1	algI	acetylase	2	3	3	5	11	12	3	10	0,27789	-1,8474	0,0851
	ARG85613.1	algI	acetylase	3	4	3	8	19	18	3	15	0,21862	-2,1935	0,0792
	ARG85613.1	algI	acetylase	8	13	12	90	145	137	11	124	0,08649	-3,5313	0,0219
	ARG85613.1	algI	acetylase	7	3	6	92	196	193	6	160	0,03469	-4,8495	0,0457
	ARG85613.1	algI	acetylase	3	3	3	51	68	85	3	68	0,04488	-4,4779	0,0222
	BAR66409.1	algA	PMI-GMP	7	7	7	14	24	21	7	19	0,34601	-1,5311	0,0437
	ARG90425.1	algC	PMM	10	9	11	27	40	37	10	35	0,28146	-1,8290	0,0215

Green, indicates upregulation; light red, indicates downregulation; yellow, indicates p-value <0.05

5.0 RESULTS: BIOMARKER IDENTIFICATION OF *LEPTOSPIRA SPP.*

5.1 Genomic Library Construction

5.1.1 Construction of genomic DNA libraries

Genomic DNA of *Leptospira spp*. was fragmented by sonication and cloned into pHORF3, resulting in libraries with 3.0 x 10^7 (Malaysian local strains library, i.e. library I) and 2.2 x 10^7 (WHO strains library, i.e. library II) independent clones. The insert rates and sizes were analyzed by colony PCR and sequencing. Thirteen randomly picked clones of each library were used for colony PCR and an additional seven per library were picked for sequencing. The average insert size was 250 bp and 290 bp for library I and library II, respectively. All libraries had insert rates of more than 80%.

5.1.2 Library packaging and analysis

Both genomic libraries were then packaged with Hyperphage for the selection of open reading frames. The packaged libraries were checked for the number of inserts by colony PCR and for correct in-frame inserts by sequencing. The cloned libraries had an in-frame insert ratio of approximately 50 to 55% and phage titers of 2.9 x10^{10} and 4.0 x 10^{10} CFU / mL. The average DNA fragments were shorter compared to the initial size, i.e. 160 bp and 120 bp for library I and II, respectively.

5.2 Selection of Immunogenic Oligopeptide Phage Pannning

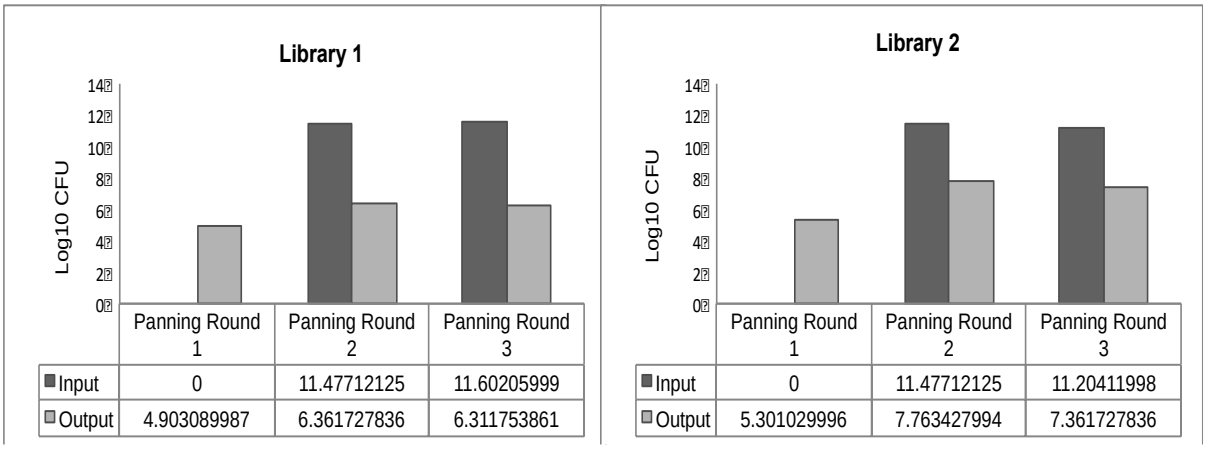
5.2.1 Panning

The pooled sera from patients with acute leptospirosis were used as polyclonal antibody source in the panning rounds and screening. Individual clones were picked after the second and third panning rounds during which interaction partners with low affinities were removed, thus selecting the best interaction partners. The results of this panning are summarized in Figure B1.

From both libraries combined, this resulted in the selection of 92 oligopeptide phage clones to be screened by ELISA. Of these, 35 had signals two-fold higher than the negative control and were analyzed further by sequencing (Figure B2). Eighteen of these 35 clones were identified as unique sequences and matched *Leptospira spp*. sequences according to BLAST analysis Altschul et al., 1997). Their encoded amino

acid sequences were translated using Expasy (Artimo et al., 2012) and the corresponding *Leptospira spp*. proteins identified using BLAST (Table B2)

Figure B1. Evolution of phage titers during 3 panning rounds with two *Leptospira spp.* genomic libraries**.** Library 1, Malaysian strains. Library 2, WHO reference strains.



	Panning Round 1	Panning Round 2	Panning Round 3
Input	0	11.47712125	11.60205999
Output	4.903089987	6.361727836	6.311753861

	Panning Round 1	Panning Round 2	Panning Round 3
Input	0	11.47712125	11.20411998
Output	5.301029996	7.763427994	7.361727836

Table B2. BLAST results of the 18 unique oligopeptide phage clones. Summary of peptides displayed by oligopeptide phage clones screened in capture ELISA. Information is given on the peptides identified and the frequency at which they were found.

No	No. Repeat	Identification	Locus	Peptide Position
1	5	Hypothetical protein LEP1GSC042_0155 [*Leptospira kirschneri* serovar Bim str. PUO 1247]	WP_042710668	35-52
2	4	Hypothetical protein LEP1GSC112_0402, partial *[Leptospira interrogans* serovar Pomona str. UT364]	EMO00657	17046
3	3	hypothetical protein LEP1GSC124_0783, partial [*Leptospira interrogans* serovar Pyrogenes str. 200701872]	WP_061236626	274-322
4	3	Hypothetical protein [*Leptospira interrogans*]	WP_017853878	131-177
5	3	Peptidase, M48 domain protein [*Leptospira interrogans* serovar Canicola]	WP_082285734	514-554
6	2	Hypothetical protein [*Leptospira kirschneri*]	WP_082292950	1231-1243
7	2	Permease [*Leptospira alexanderi*]	WP_078128696	152-167
8	2	Glucose-1-phosphate cytidylyltransferase *[Leptospira kirschneri*]	WP_004763753	87-126
9	2	AraC family transcriptional regulator [*Leptospira fainei*]	WP_016550011	239-266
10	1	PBP1A family penicillin-binding protein [*Leptospira interrogans*]	WP_061286131	685-724
11	1	Glycosyltransferase [*Leptospira sp.* P2653]	WP_083867789	13-29
12	1	DUF541 domain-containing protein [*Leptospira kirschneri*]	WP_016560806	51-67
13	1	Arylsulfatase *[Leptospira fainei*]	WP_016550521	530-550
14	1	Sterol desaturase family protein [*Leptospira*]	WP_100784316	242-260
15	1	Hypothetical protein LEP1GSC150_4200 [*Leptospira interrogans* serovar Copenhageni str. LT2050]	EMG21699	129-360
16	1	Outer membrane protein, TIGR04327 family [*Leptospira fainei*]	WP_016551048	239-259
17	1	Flagellar filament outer layer protein Flaa [*Leptospira mayottensis*]	WP_002746704	29-100
18	1	MFS transporter, partial *[Leptospira interrogans*]	WP_025176752	0-80

After three rounds of panning, screening ELISA were performed on phage clones derived from both libraries. Monoclonal phage clones displaying oligopeptides were captured in the wells of a microtiter plate using a monoclonal anti-M13 antibody and screened for reactivity with pooled sera from 18 patients with acute leptospirosis or from 10 healthy donors and detected with a goat anti-human HRP conjugate (Figure B2).

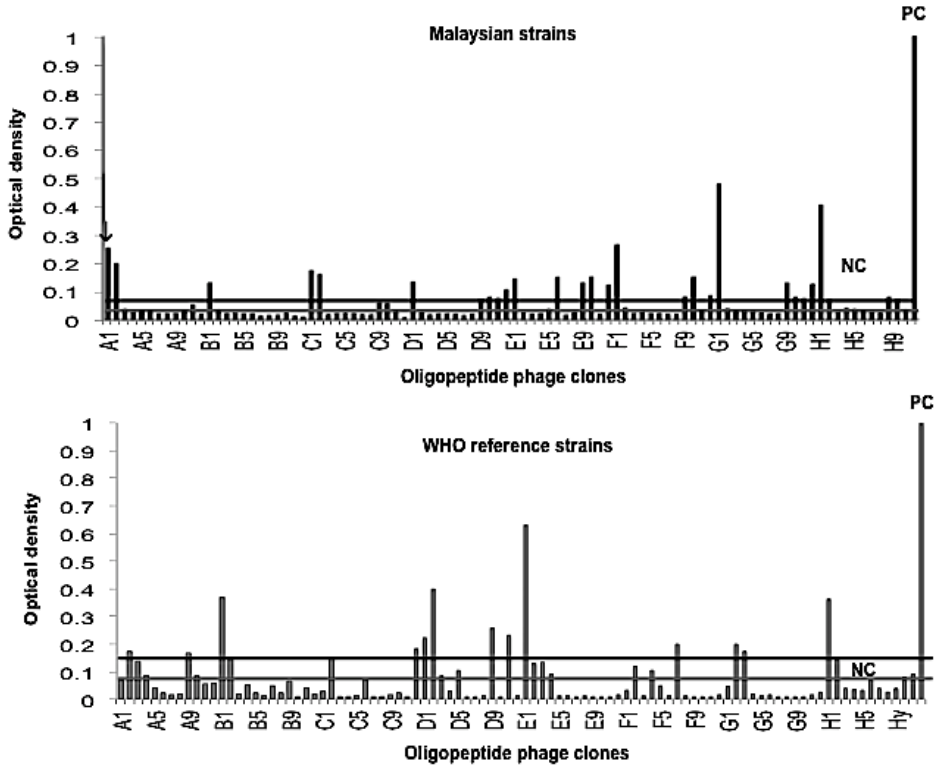


Figure B2. Screening ELISA of oligopeptide phage clones. (A) Malaysian strain (B) WHO reference strain. Anti-lysozyme antibody clone DM321-F11 was used as negative control/background (H6). Captured DM321-F11 phage particles were detected with an anti-M13 HRP conjugate as positive control for phage production and immobilization (H12). Letter-number combinations on x-axis represent the microtiter plate wells. PC, positive control; NC, negative control.

5.2.2 Production of single oligopeptide phage clones

The BLAST results of 18 unique oligopeptide phage clones corresponded to amino acid sequences of peptides consisting of 13-80 amino acids. Six clones corresponded to hypothetical proteins of unknown function; of these, the most frequent one was LEP1GSC042_0155, which was identical in five clones.

Altogether, 18/35 clones (51%) corresponded to hypothetical proteins, which agrees with the report that around 40% of genes of *L*. *interrogans, borgpetersenii and biflexia* encode proteins of unknown function (Adler and de la Peña Moctezuma, 2010). The peptide fragments of known proteins included various enzymes, transporter proteins, and outer membrane protein.

5.3 Identification of Oligopeptide Phage Clones

5.3.1 Selection of immunogenic Leptospira spp. protein fragments

The selected 18 oligopeptide phage clones produced as monoclonal phage were tested in ELISA using serial dilutions of pooled sera reactive with Malaysian strains, WHO strains and from healthy donors as described above (Figure B3).

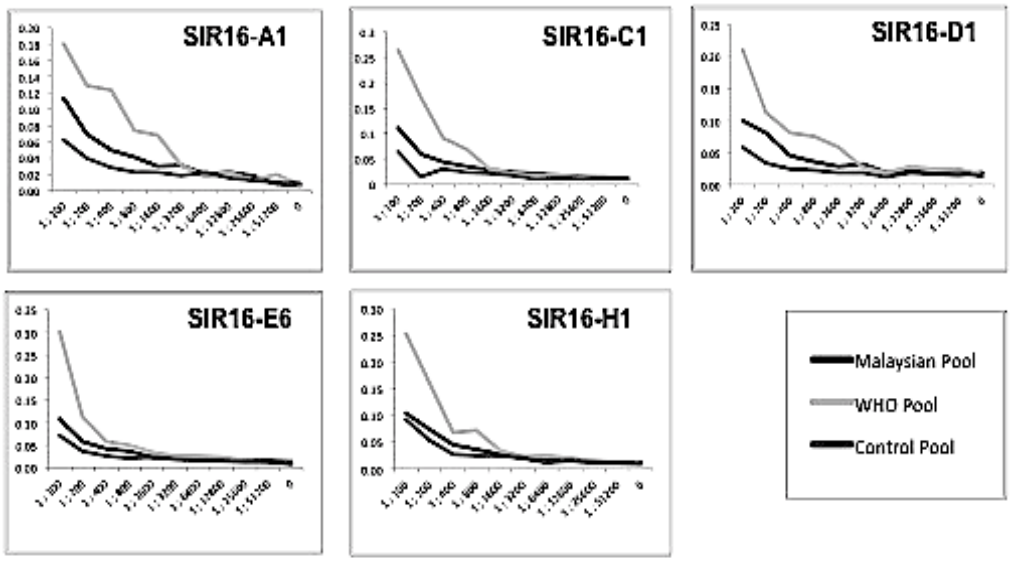


Figure B3. Titration ELISA to assess immunoreactivity of five selected oligopeptide phage clones**.** Pools from Malaysian patients sera with acute leptospirosis, which were classified by MAT according to their reactivity with circulating Malaysian strains (Malaysian Pool, n=8), and WHO reference strains (WHO Pool, n=10). A pool of non-reactive sera from healthy donors from Germany (Control Pool, n=8) was used as control.

The peptides of the five most promising phage clones were then synthesized as biotinylated peptides and tested for immunoreactivity against 16 positive and 16 negative individual sera which were not included in the aforementioned two serum pools. The two most abundant proteins in pathogenic *Leptospira spp*., i.e. the *lipL32*

(Haake, 2000) and *loa22* (Koizumi et al., 2003) gene products, were used as controls.

5.4 Validation of Immunogenic Peptide

5.4.1 ELISA for validation of immunogenic peptides

Based on the signal to noise ratio in the screening ELISA and the BLAST analysis, five peptides (SIR16-A1, SIR16-C1, SIR16-D1, SIR16-E6, SIR16-H1, as summarized in Table B3) were selected for further validation. Of note, all five selected peptides were derived from the Malaysian genomic library.

Table B3. Details of seroreactive peptides selected by titration ELISA. Titration ELISA was performed on oligopeptide phage clones using pooled sera reactive with Malaysian strains, WHO strains, and from healthy donors as detailed in Figure B3.

No. in Table 2	Clone	Blastn	Full length Protein	Peptide Sequence
10	SIR16-A1	*Leptospira interrogans* serovar Copenhageni strain FDAARGOS_203 chromosome	PBP1A family penicillin-binding protein	KSSISLGRGQA ASVLAVPIWGR MYNRFYGGQN YPSFGED
7	SIR16-C1	*Leptospira interrogans* serovar Hardjo-prajitno strain Hardjoprajitno chromosome 2	Permease	FGWNRDHFYL DGFFGSY
8	SIR16-D1	*Leptospira interrogans* serovar Manilae strain UP-MMC-NIID HP chromosome 1	Glucose-1-phosphate cytidylyltransferase	NPTAEDWEVDL VDTGALTMTGG RLLRLKDQLSK ETFMVTY
14	SIR16-E6	*Leptospira interrogans* serovar Hardjo-prajitno strain Hardjoprajitno chromosome 1	Sterol desaturase	EEPIYGLTKPVT TFDPVYT
4	SIR16-H1	*Leptospira interrogans* serovar Naam str. Naam ctg7180000007595	Hypothetical protein	EFSKTIVEKANQ FWMMVRGEGA YSKPTRISQFSI QGLMREEDVLK TS

A validation test was carried out by a titration ELISA (1:200) using individual sera from 16 patients with acute leptospirosis and 16 healthy donors (Figure 2). Two recombinant leptospiral reference proteins (rLipL32 and rLoa22) were included for comparison. The results were also compared to *Leptospira* culture supernatant antigen used for in-house ELISA. Two peptides (SIR16-D1, SIR16-H1), the two

reference proteins, and the leptospiral antigen reacted significantly more strongly with the patient sera than with the control sera, indicating high immunoreactivity (Figure B4-A). In addition, signal strength was comparable among the two peptides and reference proteins.

ROC curve analysis revealed that the two peptides (SIR16-H1 and SIR16-D1) had AUCs close to or greater than 0.8 and thus demonstrated potential as diagnostic biomarkers to differentiate between acute leptospirosis and controls (Figure B4-B).

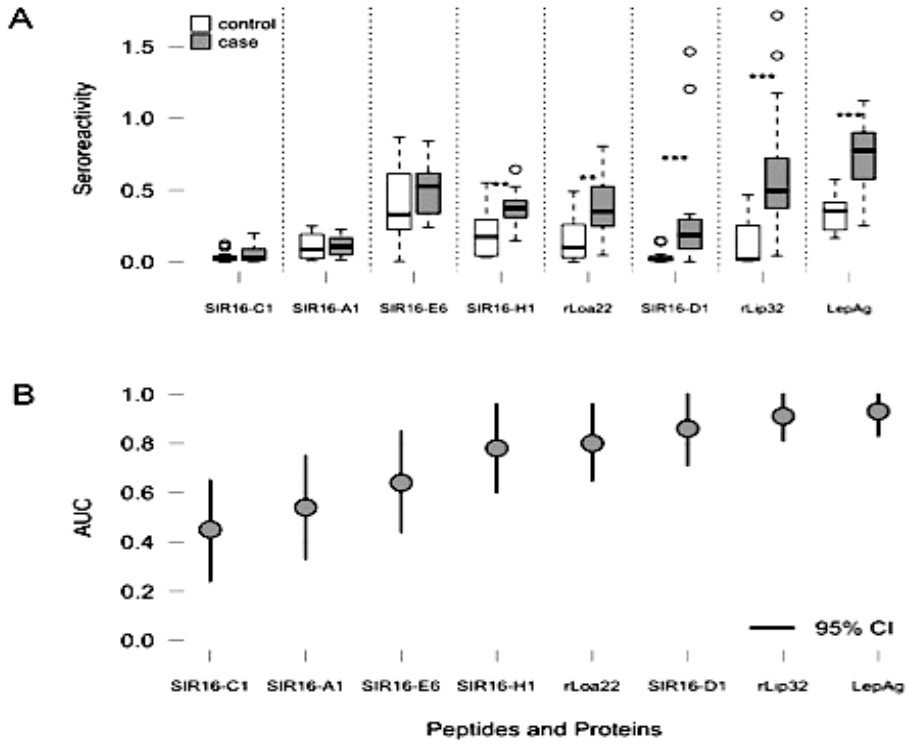


Figure B4 (A) Box plot showing seroreactivity of five peptides, two reference proteins and *Leptospira* culture supernatant antigen against sera from healthy controls of European ethnicity and Malaysian patients with acute leptospirosis (n=16 each). Serum dilution,1:200. **(B**) Area under the ROC curve (AUC) values of the five peptides and the three references antigens (rLoa22, rLip32 and *Leptospira* culture antigens). Peptides SIR16-H1 and SIR16-D1 demonstrate discriminatory ability comparable to the reference proteins. The vertical lines delineate 95% confidence intervals (CI). Stars indicate statistically significant difference between case and control groups; **p<0.01, ***p<0.001.

5.5 Validation of Novel and Established Biomarkers for *Leptospira spp.*

Logistic regression was then used to evaluate whether combinations of the peptides with each other and / or with the two recombinant reference proteins would improve classification. First, using a simulation by resampling from the existing dataset (Hosmer and Lemeshow, 1989) we made a posteriori power analysis for the logistic regression model, which includes four predictors and has the highest AUC. For the 32 samples used in the analysis, we detected a power of 98.9% at a significance level 0.001. Among the 5 peptides, a combination with better classification than the best single peptide could not be identified (Table B4, row 3). This was unexpected, as the reactivity with the individual sera correlated only weakly among the peptides, suggesting that there would be diagnostic synergy (Figure B5). However, when combining the 5 peptides with the two reference proteins, a classifier consisting of the two best peptides, rLipL32 and the peptide SIR16-A1 was identified that possessed near perfect (AUC, 0.98) discrimination between patient and control sera (Table B4, rows 4 and 6). When combining only the two best peptides with the reference proteins, classification was somewhat less accurate (AUC 0.93 vs. 0.98; compare rows 12 and 14 with 4 and 6), demonstrating the added value of including peptide SIR16-A1.

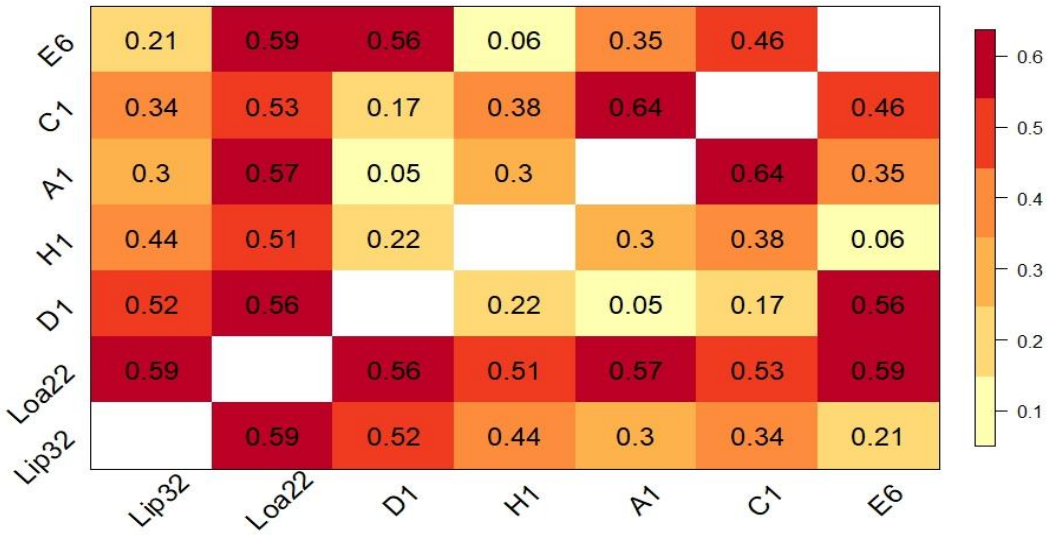


Figure B5. Correlations among the five selected peptides and the two reference proteins in terms of immunoreactivity (assessed by ELISA) with the patients and control sera. Data were based on the results shown in Figure B4. Values correspond to Pearson correlation coefficient. Only low correlation is detected between the two best peptides, SIR16-D1 and SIR16-H1.

Inspection of the data then revealed that SIR16-A1 had a lower reactivity with two of the control sera, likely explaining the observed improved classification in this multiple regression model. However, when combining the 5 peptides with the two reference proteins, a classifier consisting of the two best peptides, Lip32 and the peptide SIR16-A1 was identified that possessed near perfect (AUC, 0.98) discrimination between patient and control sera (Table B4, rows 4 and 6). When combining only the two best peptides with the reference proteins, classification was somewhat less accurate (AUC 0.93 vs. 0.98; compare rows 12 and 14 with 4 and 6), demonstrating the added value of including peptide SIR16-A1.

Table B4. Diagnostic performance parameters of single peptides, peptide combinations, and combinations of peptides with reference proteins. Results displayed based on binary comparison of leptospirosis versus healthy donors. Data are based on trade-off values in the ROC curve for each classifier.

No.	Set	Best combination	AUC (95% CI)	P value*	Accuracy^	Sensitivity	Specificity	PPV	NPV
1	Individual peptide	D1	0.85 (0.70-1.0)	3.03e-04	0.844	0.875	0.813	0.824	0.867
2	Individual peptide	H1	0.78 (0.59-0.96)	5.47e-03	0.812	0.938	0.688	0.750	0.917
3	5 peptides	D1	0.85 (0.7-0.98)	3.02e-04	0.844	0.875	0.813	0.824	0.867
4	5 peptides + rLipL32	A1+D1+H1+Lip32	0.98 (0.77-1.0)	1.83e-08	0.969	1.0	0.938	0.941	1.000
5	5 peptides + rLoa22	A1+ Loa22	0.96 (0.66-1.0)	1.86e-07	0.938	0.875	1.00	1.00	0.889
6	5 peptides + rLipL32+ rLoa22	A1+D1+H1+Lip32	0.98 (0.77-1.0)	1.83e-08	0.969	1.0	0.938	0.941	1.000
7	D1+H1	D1	0.85 (0.70-0.98)	3.02e-04	0.844	0.875	0.813	0.824	0.867
8	D1 + rLipL32	LipL32	0.91 (0.81-1.0)	1.67e-05	0.844	0.938	0.750	0.789	0.923
8	D1 + rLoa22	D1	0.85 (0.70-0.98)	3.02e-04	0.844	0.875	0.813	0.824	0.867
10	H1 + rLipL32	LipL32	0.91 (0.81-1.0)	1.67e-05	0.844	0.938	0.750	0.789	0.923
11	H1+ rLoa22	Loa22	0.80 (0.65-0.96)	2.42e-03	0.781	0.813	0.750	0.765	0.800
12	D1 + H1+rLipL32	D1 + H1+rLipL32	0.93 (0.82-1.0)	3.54e-06	0.938	0.875	1.00	1.00	0.889
13	D1 + H1+ rLoa22	D1 + H1+ rLoa22	0.85 (0,69 -1.0)	2.5e-04	0.844	0.687	1.00	1.00	0.762
14	D1 + H1+ rLipL32+rLoa22	D1 + H1+ rLoa22+rLipL32	0.93 (0.82-1)	2.66e-06	0.938	0.875	1.00	1.00	0.889

AUC, area under ROC curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value.
*Uncorrected asymptotic P values.
^ sensitivity + specificity / 2

6.0 DISCUSSIONS: WHOLE GENOME ANALYSES OF *LEPTOSPIRA SPP.*

6.1. Clinical Cases and Animal Model

In this study, strain Langkawi and strain 782 (the two most severe strains) have symptoms of Weil's disease, an icteric, severe form leptospirosis (Haake and Levett, 2015). Jaundice and subconjunctival hemorrhage are significant risk factors of severe leptospirosis ($p \leq 0.05$) (Tubiana et al., 2013, Marasinghe et al., 2017).

Some laboratory investigations of the six strains were significantly associated with severe leptospirosis i.e. hemoglobin counts <12 g/dL in strain 782 and strain 1548, and blood cell counts > 11 x 10^9 cells / L in strain Langkawi, strain 782 and strain 1530, platelet counts <100 x 10^9 / L in strain 782, serum creatinine >120 µmol/L in strain 898, strain 782 and strain Langkawi (Marasinghe et al., 2017). From the clinical presentations and laboratory findings, we can already hypothesize that both strain 782 and strain Langkawi cause a severe form of leptospirosis.

Despite the severe presentation of strain Langkawi, the patient was only admitted in the ICU for 2.5 days for resuscitation and mechanical ventilation, while strain 782 patient who received resuscitation and mechanical ventilation management was admitted to the ICU for 4 days. The other two cases stayed for 1-2 days duration. This was most likely due the difference in the policy or terms and regulations of discharging a patient from the ICU of the Dutch healthcare centers compared to that of the Malaysian. Nevertheless, the Dutch patient was hospitalized the longest period i.e. 11 days, which is consistent with the complicated disease state and management.

Among all six strains, the shortest incubation period was strain Langkawi i.e. less than 7 days (Wagenaar et al., 2005) , this was followed by strain 782 i.e. 1 week, strain 1489 (2 weeks) and strain 1530d (3 weeks). The incubation periods of the other two strains could not be determined. Remarkably, the lesser the incubation periods is associated with more severe illnesses in these patients.

As we already know, the first week following exposure to leptospirosis is the period of when leptospirae are circulating in the blood (leptospiraemia), which is then

followed by the increase of IgM antibody level in the circulation. The period when the patient of strain Langkawi was admitted to the hospital, there were still some leptospirae present in the blood although at the same time, IgM antibody have started to rise. This may explain the success in getting a positive culture (which turned positive after 4 months) from blood taken during early admission, and the ELISA test to be positive with a titer of 1:160 performed on day 2 of admission. The ELISA readings later increased to 1 : 5,120 on day 14 together with a strong positive MAT test with *L. interrogans* serogroups Cynopteri and Icterohaemorrhagiae. On the other hand, although leptospirae culture of five Malaysian cases resulted in positive isolation (after an average of 4 to 6 weeks of incubation), *Leptospira* IgM antibody ELISA was not detected in all cases. There are two hypotheses to explain this situation, either the IgM antibody all patients were still not raised above the IgM detection level or the IgM antibody test used in HTAR Klang hospital was not as sensitive as the one used in the Dutch Hospital for strain Langkawi patient. As all *Leptospira* IgM ELISA tests were negative, none of the cases were tested for MAT.

As there is no issue of antibiotic resistance in leptospirosis, the selection of antibiotics was quite broad in range, to most antibiotics (beta-lactams, cephalosporins, aminoglycosides, macrolides) but not to chloramphenicol, vancomycin, rifampicin and metronidazole, spectinomycin, amphotericin B, 5-fluorouracil, fosfomycin, trimethoprim, sulfamethoxazole, neomycin, (Faine, 1999; Kobayashi, 2001; Chakraborty et al., 2010). All Malaysian patient were treated with combination of Doxycycline and Ceftriaxone which was the second line recommended regime by Malaysian Antibiotic Guideline 2014. This regime was selected over Doxycycline and Penicillin G due to the once-daily administration and the extended spectrum of ceftriaxone against bacteria provide additional benefits (Panaphut et al., 2003).

The severity of a disease can be due to pathogen factors, (i.e. type of serovar / species and inoculum size) as well as host factors (patients' immunological response) that contribute to this heterogeneity in disease effects. In this case study, all six patients had various histories of exposure, incubation periods and clinical presentations. In order to determine which *L. interrogans* strains were the most and least virulent, we eliminated the influence of the immunological state of the patients

by examining the effect by infecting the strains at a standard concentration (i.e. 10^8 leptospires/mL) into animal models. The severity of disease was compared in clinical settings as well in the animal models. Amazingly, the severity of diseases observed in the patients were well correlated in animal models, which suggests that the disease spectrum in these patients were predominantly determined by the strain virulence itself and minimally influence by the immunology status of the hosts.

Although the guinea pigs infected with strain Langkawi survived for two or three days longer than those infected with strain 782, they had much severe disease complications, as evidenced by histopathology findings. In this study, histopathology findings are taken as the confirmatory test for the effects and complications of the disease.

Intra-alveolar haemorrhage is known to be a severe complication of leptospirosis, and associated with a poor prognosis (Gulati et al., 2012). Here, we not only assessed the severity of the hemorrhage in the lungs, but we also measured the area of haemorrhages in the lungs of each guinea pigs and compared differences among infection using statistical analysis. As shown in the 'Results' chapter, strain Langkawi had the most severe and widest area of intra-alveolar haemorrhage, while strain 1530d had the mildest and most limited area of intra-alveolar haemorrhage in one guinea pig out of three. This was followed by strain 1548 with two out of three guinea pigs had mild intra-alveolar haemorrhage. Three strains i.e. Langkawi, 782 and 898 showed involvement of the kidneys with presence of leptospirosis pathognomonic findings in the kidneys i.e. interstitial nephritis and/or pyelitis and/or tubular degeneration. Strain 782 showed the moderate severity in kidney findings while strain Langkawi and 898 just caused mild changes in the kidneys.

The interpretation of clinical chemistry laboratory results of the seven guinea pig groups is quite challenging, as there is no standard reference range for any of the parameters. The biochemistry tests suggested that there was derangement of kidney function and increased inflammation in some of the guinea pigs and raised bilirubin serum in strain Langkawi only. The second laboratory testing performed on the most and least virulent strains confirmed that strain Langkawi caused a higher level of infection and inflammatory process besides causing hyperbilirubineamia and acute

kidney injury. Both strain Langkawi and 1530d showed deranged liver enzymes. However, this was more prominent in stain 1530d. Detection of leptospires by qPCR not only validate the diagnosis, but also verifies the level of disease severity of the strains which was well correlated with the leptospires burden.

6.2 Typing & Phylogenetic Analysis

At present, the most common molecular typing method used to establish strain relatedness in *Leptospira spp*. is MLST due to its reproducibility and high discriminatory power (Chiani et. al, 2015). At present, there are three schemes being developed for *Leptospira spp*. MLST, using different sets of six or seven housekeeping genes. This fact demonstrates that efforts are being made to have a comprehensive, precise and feasible molecular typing for this fastidious, slow-growing yet highly variant bacteria.

In this study, we have initiated whole genome sequence typing using two different approaches, i.e, gene by gene locus typing and SNP calling. As *L. interrogans* is known to be highly diverse with a variety of serovars and strains and the pan-genome of *Leptospira* spp. is open (i.e., the number of new genes is still increasing with the number of additionally sequenced strains (Fouts et al., 2016), an ad-hoc cgMLST scheme was developed using strains isolated in Malaysia. The ad-hoc cgMLST and MLST constructed from 29 strains showed congruent findings for strains with similar ST in MLST.

cgMLST relies on the availability of complete, accurately sequenced genomes for the generation of the typing schemes. Inclusion of coding sequences only decreases the number of sites typed in the analysis, but at the same time it facilitates standardization and reproducibility of the analyses as it focuses on a predefined set of genes. In WGS analysis the quality of the reads as well as the assembly play an important role in achieving reliable cgMLST results (Janowicz et al., 2018).

SNP analysis potentially has the highest discriminatory power among the typing methods, as polymorphisms can be discovered in both coding and noncoding regions of the genome. However, the choice of a reference genome can significantly

influence the number of identified SNPs and the accuracy of the reconstructed phylogenetic relationships (Pightling et. al, 2014).

In this study we found that SNP analysis (using the ParSNP tool) differentiated the strains better than cgMLST, most likely due to better coverage of the core genome in SNP analysis (77.5%) compared to cgMLST (57.7%). The core genome of cgMLST (2726 genes) and parSNP (2841 sequence clusters) was lower compared to core genome in analysis with protein orthology (3271 core coding sequences) as there is different formula being used in cgMLST and SNP analysis.

The distances between genomes using the SNP calling method were much lower compared to cgMLST due to different calculators being used. In parSNP the strains were distanced by the relative number of SNPs present in certain sequence regions, whereas in cgMLST the strains were distanced by the number of gene differences (alleles). In the latter case, even one single nucleotide difference in a certain gene leads to a complete mismatch (100%) in comparison to SNP calculation, where one nucleotide difference in a single gene (~1000 bp) leads to only 0.01% mismatch (Treangen et al, 2014). Thus, resolution of the cgMLST method is much lower than cgMLST and already small differences such as a single mismatch or sequencing errors lead to massively increased distance values. Thus, cgMLST can only be used for intra-species typing, whereas SNP-based typing can be used for differentiation of different species as well.

In our study, we found that the gene-by-gene approach was the SNP analysis are in congruent to ST clustering in MLST. However, the topology placement of 29 *L. interrogans* strains were slightly differ as SNP analysis has higher coverage of the core genome and higher degree of discrimination between the strains. Furthermore, the branch length of the phylogenetic tree in cgMLST and parSNP was not comparable, as both analyses use different calculators to obtain branch distance.

The advantages of both methods compared to conventional *Leptospira* typing are high resolution, WGS-based typing which are able to delineate intra-species differences, i.e. strain definition at the molecular level. However, at this point of time, WGS typing might be unlikely a realistic approach for epidemiologic studies of leptospirosis outbreaks or surveillance due to its fastidious and slow growing nature.

In the advent of metagenomic studies and advances in genomic technology, the possibility of bacteria genomes to be recovered and whole genome sequenced without isolations is becoming more realistic and thus increase the chances of WGS typing for a fastidious organism like *Leptospira spp.*

6.3 Genomics Findings

Among the six strains of *L. interrogans*, there are 3271 common coding sequences referred to as the core genome. Analysis of functional annotation by COG showed that the majority of CDS in the core genome are unknown in function, which agrees with previous findings that almost half of overall *Leptospira* genome are hypothetical proteins (Adler, 2004).

The presence of variable number of plasmids and insertion sequences in the six *L. interrogans* genomes together with strong genome rearrangements suggested a high plasticity even amongst strains of the same species. This may occur through homologous recombination and horizontal gene transfer (HGT), which may result from three mechanisms: transformation, transduction and conjugation (Bryant et. al, 2012; Levin and Bergstrom, 2000; Darmon and Leach, 2014). The highest number of insertion sequences and plasmids are present in strain 1489, whereas strain 898 has the least number of insertion sequences and absence of plasmid in the genome. Furthermore, as from MAUVE we observed limited synteny within the large chromosomes. In most of the cases those result from gene recombination events between insertion sequences.

Previous comparative genomic analyses of leptospires described the presence of prophages in the *Leptospira* genomes (Ricaldi et al., 2012; Schiettekatte et al., 2018). However, in recent reports by Fouts et al., 2017 prophages were found in numerous copies in pathogenic and intermediate leptospires, compared to saprophytic leptospires. This had suggested that there were evolution of virulence in pathogenic leptospires (Fouts et al., 2016). Surprisingly in this study, all prophages found in the six genomes appear to be pseudogenes as they are incomplete and non-functional. Frameshifted genes accumulate, which have undergone disablement of gene function through point mutations, insertions and deletions resulting in

misplaced internal stop codons (Tutar, 2012)

CRISPR arrays were found in all six strains of *L. interrogans*. This reflects the depth of bacterial adaptive immune responses to foreign genetic elements to the host in the past. On the other hand, CRISPR arrays are most frequently present in the two strains of serogroup Bataviae, followed by strain of serogroup Canicola and lastly in serogroup Icterohaemorrhagiae.

The *rfb* locus in *Lepotspira spp.* encodes for LPS biosynthesis, which is the key antigen in immunity to leptospirosis. The locus was defined on the basis of the identification of proteins involved in the biosynthesis or polymerization of nucleotide sugars (Bulach et al., 2000; Kalambaheti et al.,1993). Among the six *L. interrogans* strains, the longest *rfb* locus was noted in strain 782 (95 genes) followed by strain 1530d (90 genes). The shortest *rfb* locus is present in strain 1548 (77 genes). Furthermore, frameshift mutations were postulated to take place in two strains of Bataviae serogroups (strain 1489 and 1548); evidenced by presence of consecutive transposon in the locus. Interestingly, although strain 782 and strain 1489 has the longest and most complex *rfb* locus respectively, it was less virulent compared to strain Langkawi, Of note, the *rfb* locus of serogroup Icterohaemorrhagiae are more conserved with uniform gene arrangements and minimal gene difference. Strain Langkawi possess a unique gene combination from strain 782 and 1530d as well as an addition of Galactoside O-acetyltransferase, which distinguish it from other Icterohaemorrhagiae serogroup strains. The diversity of leptospiral O-antigen resulted from antigenic changes in previous infections and environmental ecology-mediated selection pressures, as reported in *Salmonella spp.* (>2000) serotypes), (Nally 2005a; Fouts et al., 2016; Feasey et al., 2012; Wildschutte et al., 2004)

Lipid A in *L. interrogans* is known to be less endotoxic than that of the Enterobactericeae (Que-Gewirth et al., 2004). In actual fact, The GlcNAc3N with four amide-linked fatty acids in *L. interrogans* resembles the features of the environmental bacteria, which hypothesize their ancestral origin (Que-Gewirth et al., 2004; Slamati et al., 2011; Picardeau et al., 2017). Being a motile bacterium, there was an increased drive to disseminate into a new ecological niche / reservoir. Equipped with complex biological system, *Leptospira* have become a versatile bacteria that is able to live in natural environment as well as maintaining its survival

in immunoprivileged area i.e proximal renal tubules in various animal host, and later get excreted in urine (Athanazio et al., 2008). By going through this viscious cycle, leptospires undergo homologous recombination and HGT in order to enhance its survival and virulence potential.

In this study, numerous plasmids and IS were seen in each of the strain genome, most likely explained the edition number of modification and adaptation to carrier hosts previously. On the other hand, the higher the number of editions to the genome i.e., represented by number of mobile genetic elements (MGE) in *L. interrogans,* does not correlate to the higher the intensity of the virulent, which can be seen between strain Langkawi strain 1489 (lesser virulent with most MGE) and strain Langkawi (most virulent with lesser MGE). Nevertheless, MGE most likely contribute to improved adaptation as well as virulence

Until now, it is still not clear which genes of pathogenic *Leptospira spp.* are exactly responsible for their differences in virulence. Although it was found that nearly 900 genes from pathogenic strains are absence in saprophytic *Leptospira spp*. (Adler, 2011), the essential virulence factors are still blurry as they work in functional redundancy manner. For instance, the outer membrane proteins of leptospires do not only play an important role within adhesion to ECM during the invasion process, in fact they also stimulate the host immune system, which eventually leads to cytokines storms in septic shock complication (Adler, 2011; Cagliero et al., 2018).

6.4 Comparison of The Most and The Least Virulence Strain

From this study, we can deduce that in both human cases and animal models, infection of strain Langkawi showed the most severe disease complications. Clinically patient had septic shock and SPHS; laboratory results showed hematology profile derangement, elevated acute phase proteins and liver and renal function markers. Histopathology of animal models showed moderate to severe intra-alveolar haemorrhage with significantly the highest densiometric measurement of haemorrhage areas compared to infection with the other five strains. The least severe strain was strain 1530d, which showed minimal signs and symptoms of leptospirosis, less pronounced changes in laboratory findings, and only one guinea pig had mild intra-alveolar haemorrhage with lowest densiometric score.

Interestingly, both strain Langkawi and strain 1530d were from serogroup Icterohaemorrhagiae. Although some studies have suggested that *Leptospira spp*. from serogroup Icterohaemorrhagiae cause more severe disease than other serogroup (Herrmann-Storck et al., 2010; Tubiana et al., 2013), our findings in this study disagree with this hypothesis. Instead, it agrees with the observation by VInetz, 2001; Helmerhorst et al., 2012 that there is a poor correlation between particular serovars and severity of leptospirosis.

For strain Langkawi, the incubation period was the shortest compared to the rest of the strains, i.e. <1 week in the patient as well as in the animal model. For strain 1530d, the incubation period in patient was estimated to be longer as he sought treatment after ~3 weeks post exposure, which is in contrast to the animal model which showed symptoms on day seven of infection. According to previous studies, most patients experience a mild, self-limiting undifferentiated febrile illness occurring after a 1–2 weeks incubation period (Bharti et al., 2003; Plank and Dean, 2000).

The clinical features of strain Langkawi met the criteria for a diagnosis of sepsis, which suggested that development of severe leptospirosis could be highly due to dysregulated inflammation (Cagliero et al., 2018). Interestingly, data obtained from our clinical studies as well as histopathology findings in the animal model support this hypothesis. Both human and animal model were noted to suffer from pulmonary haemorrhage with background features of Weil's disease. In contrast, infection with strain 1530d was associated with an absence of known leptospirosis complications in the clinical case, and only minimal complication in the animal models.

In infection with strain Langkawi infection, both human and animal models showed that jaundice was an apparent accompanying sign. Although the liver was affected, as evidenced by raised liver enzymes and bilirubin in human and animal models, the changes seemed transient. This agrees with the study by de Rocha Medeiros et al., 2010 who reported that leptospirosis complications are associated with liver cell dysfunction but without evidence suggestive of hepatocyte loss or apoptosis seen in the histopathology findings. Clinically, plasma bilirubin concentrations are high with slightly elevated transaminase plasma concentrations. In addition, thrombocytopenia was observed during the acute stage, probably because of both diffuse intravascular coagulation and immune-mediated mechanisms (de Rocha Medeiros et al., 2010).

Urinalysis in the patient infected with the Langkawi strain revealed the presence of some erythrocytes, leukocytes, a few granular casts, and a strong reaction for protein and hemoglobin, indicating rhabdomyolysis (Wagenaar et al., 2005).

In this study, the inventory of virulence genes among the six *L. interrogans* does not differ much with respect to content or sequence homology. However, looking at the huge phenotypic differences of the infection in the two most extreme strains (ie. Langkawi and 1530d), the question arises whether differences in gene transcription could possibly play an important role in virulence. Surprisingly, of all genes associated with potential pathogenesis functions, only genes relating to chemotaxis were upregulated in strain Langkawi compared to strain 1530d. This leads to the hypothesis that the transcriptome of leptospires *in vitro* does not directly reflect its actual pathogenesis capacity *in vivo.* Within the logarithmic phase of *in vitro* cultures, the number of bacteria cells doubles at a constant, exponential rate which is proportional to the population of cell at that time, i.e. 10^8 cells /ml (Zwitering et al., 1990; Rolfe et al., 2012). At this stage, bacteria prepare themselves for an infection, at which point the competence of function associated with invasion into the host is expected to be at maximum level. According to (Mounier et al.,1997), at early logarithmic phase, *Shigella flexneri* is six time more invasive than at stationary phase, as it activates the Mxi-Spa apparatus for regulation of bacterial entry efficiency, and it secretes Ipa proteins more often at the stage of division. In our study, we found that the proportion of chemotaxis genes was transcribed more highly in strain Langkawi, i.e. two to ten-fold higher than in strain 1530d at logarithmic phase *in vitro* culture, which may explain the superiority in virulence at least in part. But, clearly, this hypothesis needsto be tested in an in-vivo transcriptomic study.

6.5 Hypothesis of Pathogenesis

The hypothesis of leptospirosis pathogenesis had been a subject of debate since the first cases of human leptospirosis were reported. By looking into the established bacterial virulence factors with significant proteins sequence homology with our six *L. interrogans* strains, we tried to improve our understanding of leptospirosis pathogenesis.

6.5.1 Pathogen entry and dissemination by offensive mechanisms

Pathogenic leptospires enter the host through breached skin /mucous membranes, and entry is facilitated by its vigorous corkscrew motility. Leptospires pass through the epidermis, dermis and connective tissue (collagens, laminin, elastin, fibronectin, vitronectin) layers, by adhering and penetrating the host tissue and alter host defends response. The invasion of leptospires mainly depends on efficient penetration and adherence to various types of membrane cells and assisted by secretion of enzymes secretion such as collagenases, sphingomyelinases and phospholipases (Charon et al., 1992; Murray et al., 2015, Kassegne et al., 2014; Narayanavari et al., 2012).

Proteins like ColA (encoding collagenase), SmcL (encoding sphingomyelinase C), Hlb and HlyB were found in the genome of the six *L. interrogans* strains. The presence of sphingomylinase C in pathogenic leptospires has been reported in Nascimento, 2004, together with a homologue of phospholipase D. Collagenase was reported to be an important virulence factor in pathogenic leptospires (Kassegne et al., 2014). Dinges et al., 2000 reported that the Hlb protein cleaves sphingomyelin into phosphocholine and ceramide whereas HlyB is a subunit of the hemolysin transport protein of *E. coli* 01570, that is involved in protein secretion by type I secretion system (Lally et al., 1997).

Invasion of leptospires into host tissues proceeds by adherence to extracellular matrix complex (ECM) and translocation into blood vessels by penetrating the cell-cell junction of vascular endothelial cadherin (Fernandes et al., 2016, Evangelista et al., 2014). The bacteria are then disseminated haematogenously to target organs, i.e. kidneys, liver and lungs. Besides ECM and endothelial cadherin, leptospires also bind to circulating polypeptides, notably plasminogen and fibrinogen. The former results in the generation of plasmin and increase in its proteolytic ability, while the latter interferes with clotting in a thrombin-catalyzed reaction and may promote hemorrhage foci. Interaction with negative regulators of the complement system may help bacteria to evade the host immune system, facilitating the invasion (Fernandes et al., 2016).

The corkscrew motility of leptospires plays a huge role in pathogenicity. Studies have shown attenuation of virulence due to mutations in motility proteins FlaA1 and LB139 (Liao et al., 2009; Lambert et al., 2012; Eshghi et al., 2014). Sequence homologies

seen in proteins that encode for flagellar motility genes resembles those of three other bacteria i.e. (1) *Helicobacter pylori* 26695 for flagellar basal body rod protein, flagellar motor switch protein G and flagellar motor switch protein; (2) *Legionella pneumophila subsp. pneumophila* for flagellar synthesis regulator, transcriptional regulator and sigma 54-dependent response regulator; (3) *P. aeruginosa* for flagellar P-ring protein precursor and flagellar biosynthesis protein.

6.5.2 Persistence of Infection by Activating Defense Mechanism

The fact that O-antigen of leptospires is twelve times less endotoxic than that of *E. coli* has been known for some time (Isogai et al., 1986). Leptospires O-antigen resembles that of gram negative bacteria, but it does not trigger TLR4 activation as the gram negative O-antigen. Instead, TLR2 is activated, which by itself is unable to mount an effective innate immune response following an infection (Chassin et al., 2009). It has been reported that lipoproteins and *S. aureus* peptidoglycan favor the induction of Th2 development by dendritic cells by triggering TLR2 (Moll, 2003; Re and Strominger, 2001). Here we found that some genes of gram positive bacteria are encoded in the *rfb* locus of leptospires i.e. Cap8EFG from *S. aureus* and NeuBC from *Streptococcus agalactiae (S. agalactiae*) respectively. It was reported that Cap8EFG provides antiphagocytosis properties in the O-antigen (Thakker et al., 1998; Cunnion et al., 2001; Cunnion et al., 2003) and NeuBC functions as binding inhibition of the activated complement factor C3b to the surface of bacteria to prevent activation of the alternative complement pathway and inhibit complement-mediated opsonophagocytosis (Marques et al., 1992; Takahashi et al., 1999; Von Hunolstein et al., 1999; Cieslewicz et al., 2001).

Besides this, leptospires have several mechanisms to survive attack by innate immune cells including macrophages or to promote survival after ingestion by phagocytic cells. KatE detoxifies H_2O_2 and protects against reactive oxygen species (ROS) including hydrogen peroxide, superoxide and hydroxyl radicals (Johnson et al., 1993; Soler-Garcia & Jersie, 2004; Wu et al., 2009; Muench et al., 2009). ClpC encodes an ATPase promoting early escape form the phagosome of macrophages (Rouquette et al., 1996; Rouquette et al., 1998; Nair et al., 2000) RecN encodes a recombinational repair protein that protects against ROS and non-oxidative killing by

neutrophils (Stohl et al., 2006; Criss et al., 2009). Transition from the environment to the host induces the up-regulation of several stress response genes (Matsunaga et al., 2007) such as catalase, KatE (Eshghi et al., 2012) and the molecular chaperone ClpB (Lourdault et al., 2011). Besides this, there are reports that leptospires survive and replicate in macrophages and induce macrophage apoptosis (Toma et al., 2011 and Li et al., 2010). Mutation of the leptospiral Hsp90 homologue, HtpG, also resulted in attenuation of virulence, but without a concomitant increase in susceptibility to oxidative stress (King et al., 2014); the basis of these differences remains unknown (Adler, 2014).

Four locus tags with sequence homology to the MtrD protein of *Neisseria gonorrhea*, which encodes resistance nodulation division (RND)-type of efflux pump were found (Shafer et al, 1998; Jerse et al., 2003). It is hypothesized to function as a supporter of growth of leptospires under hostile conditions encountered in vivo. Iron is essential for the viability of most, but not all, bacterial species. The ability to acquire iron *in vivo* is therefore a key virulence property. Pathogenic leptospires possess a heme oxygenase, HemO, which facilitates the acquisition of iron from heme and is required for virulence (Murray et al., 2008; Murray et al., 2009b). Here we found sequence homology to two proteins that function in iron uptake and assimilation in *Legionella pneumophila* i.e. FeoB (a Fe2+ transporter) and CcmF, which promotes iron assimilation and intracellular infection (Robey and Cianciotto, 2002; Viswanathan et al., 2002). Mechanisms and physiological role of Mgtb are not completely clear; however, Mgtb is a unique transport system for Mg (2+) with unusual mechanisms for mediating Mg (2+) movement through the membrane (Moncrief and Maguire, 1999).

Gram-negative bacteria are equipped with five types of double-membrane-spanning secretion systems (Costa et al., 2015), but only 2 secretion systems have been reported so far in leptospires: a Type 1 secretion system (T1SS) and a type 2 secretion system (T2SS). Although the role of T2SS in protein secretion has not been demonstrated, several components of the T2SS are encoded in the leptospiral genome. Besides T1SS and T2SS, other types of secretion systems are reported to be absent from the leptospiral genome (Abby et al., 2016; Picardeau, 2017). In this study, we found that there are sequence homologies of T2SS genes from other

bacteria in the six strains of *L. interrogans*. The proteins are GspD and GspE from *Shigella dysenteriae* and XcpS from *P. aeruginosa*, the latter of which is required for the translocation of a variety of toxins and enzymes across the outer membrane into extracellular fluid (Chapon-Hervé, 1997; Filloux et al.,1998).

It is a mystery how leptospires invade the host cells without having the type 3 secretion system (T3SS) which is a significant virulent factor for bacterial invasion and survival in host cells. Here we found a significant sequence homology of the CdsN ptotein that act as an ATPase for T3SS in *Chlamydia trachomatis* (*C. trachomatis*) in all six *L. interrogans* strains with a conserve locus tag. In *C. trachomatis* and *C. pneumoniae*, CdsN is a functional ATPase that catalyzes unfolding of proteins and the secretion of effector proteins through the injectisome and interacts with a putative chaperone, Cpn0706, and with CopN, the putative plug and effector protein, which is suggestive of a functional T3SS system in *C. pneumoniae* (Stone et al., 2008). However, proteins with sequences similar to Cpn0706 and CopN have not been found in *L. interrogans*.

It is known that biofilms production by leptospires facilitates their persistence in the environment. It is also postulated that biofilm biosynthesis *in vivo* play a major role in promoting leptospiral colonization in proximal renal tubule of carrier animals and protect them from host defends mechanisms (Treuba et al, 2004 and Ristow et. al, 2008).

From the result of protein homology of alginate biosynthesis, the inner membrane and periplasmic proteins were found to be present in all six genomes but not the transmembrane proteins. The absence of proteins particularly associated with alginate export (AlgK and AlgE) suggested there might be a different gene pertaining to these functions or there are probably unspecific transport mechanisms used for alginate export through the cell wall. Figure A13 shows a summary of protein homology associated with alginate biosynthesis, regulatory and genotypic switching in reference to that of *P. aeruginosa* PA01.

Genes for alginate biosynthesis in *L. interrogans* were scattered across chromosome 1 and were not arranged in a gene cluster as in *P. aeruginosa*. Some genes only resulted in very low e-values on alignment blast analysis (i.e. AlgX), while some

genes might have different sequence homology (e.g. Alg14, AlgZ). There is also a possibility that these genes might have undergone frameshift mutations due to high passage number. All product names in locus tags were the same for each gene, suggesting that they are conserved protein. The hypotheses of alginate biosynthesis mechanism in *L. interrogans* is displayed in Figure A13.

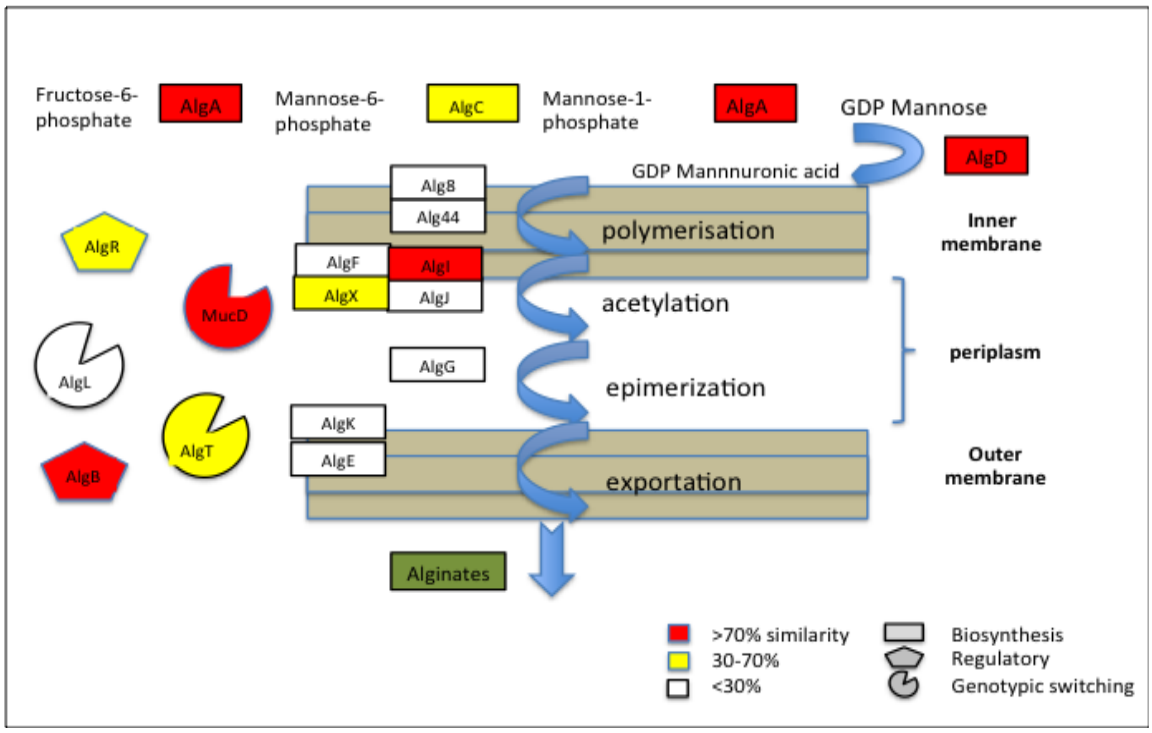


Figure A13. Hypothesis of alginate biosynthesis, regulatory and genotypic switching of *L. interrogans* in reference to alginates biosynthesis proteins of *P. aeruginosa* PA01. Alginate biosynthesis is a mechanism for survival of leptospires in environment. Biosynthesis of Alginates is initiated by conversion of Fructose-6-phosphate to GDP Mannuronic acid in the cytoplasm catalyzed by gene products of *algA*, *algC and algD*. Further conversion i.e. polimerisation (*alg8* and *alg44*i), acetylation (*algI, algJ, algF and algX*), epimerization (*algG*) and exportation (*algK and algE*) takes places. Protein homology was shown to be high (red) for proteins involved in alginate biosynthesis process in the cytoplasm (*algA* and *algD*) and inner membranes (*algI*). There is lack of sequence homology between *L. interrogans* and *P. aeruginosa* for alginate enzymes action at polymerization (inner membrane), epimerization (periplasmic) and exportation (outer membrane) suggesting different enzymes or mechanisms being used.

6.5.3 Damage to Host Tissue by Immune Reaction

During an infection, leptospires express species specific Microbial Pathogen-Associated Molecular Patterns (PAMPs) and stimulate the innate immune system of the host via Pattern Recognition Receptors (PRRs). The interactions of PAMPs with PRRs mainly take place at surface of Toll-like receptors (TLRs) and nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs) (Akira et al., 2006). The PAMPs/PRR association initiates an inflammatory cascade by activating multiple intracellular signaling pathways, that eventually regulate the expression of cytokines, prostaglandins (PGs), and nitric oxide (NO) (Schroder and Tschopp, 2010; Tisoncik et al., 2012; Turner et al., 2014; Werts, 2017).

In normal infection events, the cytokines and cytokine regulators work in a well-regulated manner to clear the pathogens without inducing tissue or organ damage. However, in a severe leptospirosis, the inflammatory response is excessive with prolonged, persistent and imbalanced activity of pro-inflammatory cytokines, i.e. IL-1β, TNF-α, IL-6, which trigger a scenario known as 'cytokine storm'. In this tragic scenario, anti-inflammatory cytokines are secreted massively but their effect is being paralyzed, which eventually can lead to sepsis and associated organ failures (Tisoncik et. al, 2012; Zhao et. al, 2015). In fact, sepsis is now defined as a life-threatening organ dysfunction caused by a dysregulated host response to an infection and is referred to as a "cytokine storm-induced syndrome" (Chousterman et al., 2017).

High concentrations of pro-inflammatory IL-6 produced by immune and non-immune cells are induced by TNF- α and IL-1β. It is also an indicator of septic shock and correlates positively with leptospirosis severity and SPHS (Reis et al., 2013; Schulte et al., 2013; Papa and Kotrotsiou, 2015). It is also reported that TNF-α, IL-1β, and IL-6 can activate the coagulation system, especially in endotoxemic models (Schulte et al., 2013). Serum levels of the pro-inflammatory cytokine IL-6, the chemokine IL-8, and the anti-inflammatory cytokine IL-10 were significantly higher among leptospirosis patients with SPHS than without SPHS (Reis et al., 2013; Mikulski et al., 2015; Chirathaworn et al., 2016). A hypothesis of leptospirosis pathogenesis in *L. interrogans* infection is displayed in Figure A14.

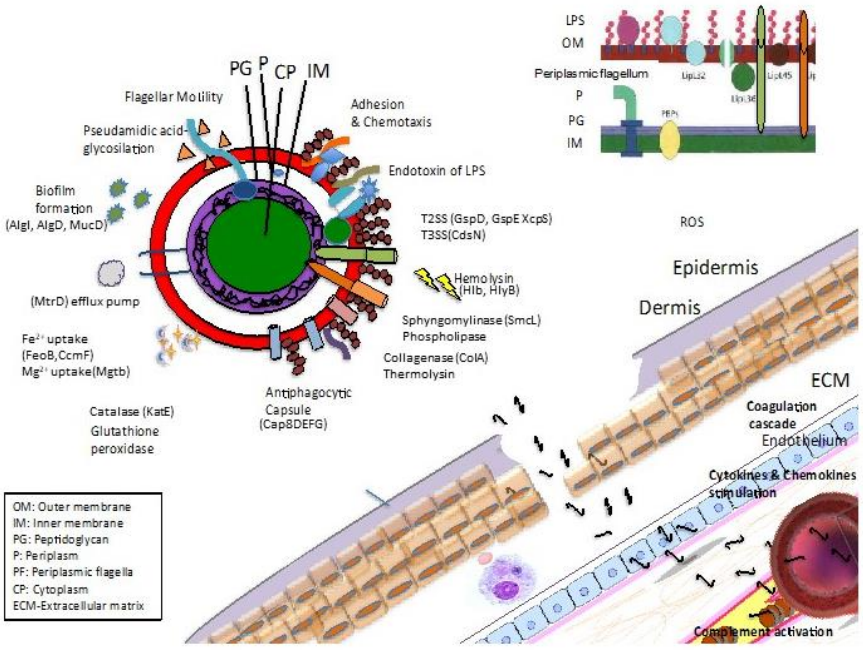


Figure A14. Hypothesis of *L. interrogans* pathogenesis. Motility-associated proteins are the earliest virulence factors. Through vigorous corkscrew movement, pathogenic leptospires penetrate into breached human skin and mucosae. Facilitated by adhesion proteins and lipoproteins on the surface of the body, as well as several types of endotoxin (sphingomyelinase, collagenase, phospholipase) and proteases (thermolysins), leptospirae penetrate and invade by adhering to tissue and degrading the membranes of various connective tissues. Oxidative stress proteins such as glutathione peroxidase and catalase *katA* and antiphagocytic capsule *cap8EFG* are defensive mechanisms against host phagocyte reactive oxygen species. The *mtrD* efflux pump mediates resistance to structurally diverse hydrophobic agents through an energy-dependent efflux process. The chemotaxin proteins direct movement toward the heme, an important *in-vivo* iron source, and the pathogen thus translocates through the vascular endothelial cadherin to make its way into the blood circulation. The mobilization of iron from heme is catalyzed by heme oxygenase *hemO,* and iron binding, transport and assimilation are facillitated by *ccmF* cytochrome c heme lyase and Fe^{2+} transporter *feoB*. Leptospires are disseminated to other organs via blood circulation. In response to leptospires, inflammatory cytokines and circulating leukocytes are stimulated. Other host molecules (including complement proteins, thrombin, fibrinogen and plasminogen) that facilitate immune evasion and bacterial dissemination are activated to contain the infection and regulate inflammatory reaction.

7.0 DISCUSSIONS: BIOMARKER IDENTIFICATION OF *LEPTOSPIRA SPP.*

ORFeome phage display has been proven to be a successful method for selection of immunogenic peptides to be used for diagnostic purposes. It has previously been used successfully to identify novel biomarkers from *Salmonella typhimurium*, *Mycoplasma pneumoniae* and *Neisseria gonorrhea* (Naseem et al., 2010; Kügler et al., 2008; Meyer et al., 2012; Connor et al., 2012). Regarding *Leptospira*, it has been used successfully to identify host proteins that interact with LipL32 (Chaemchuen et al., 2011), LigB protein acting as heparin binding protein (Ching et al., 2012), selection of protein with adhesion activity (Lima et al., 2013) and mimotopes from monoclonal antibodies specific to *Leptospira spp*. (Tungtrakanpoung et al., 2006).
Our study is the first to use phage display to identify immunogenic *Leptospira* antigens from *Leptospira spp.* genomes.

The findings of this study are of particular interest in that they indicate that ORFeome phage display can be used to identify novel peptides for development of leptospiral diagnostics, an approach that promises to be superior to protein or cell extract-based methods for use in tropical and resource-limited settings for the reasons outlined in the Introduction. For instance, an ELISA with directly immobilized peptide would constitute a simple peptide-based point-of-care diagnostics for resource-limited settings (Sagna et al., 2013; Dubois et al., 2012; Dama et al., 2013; Abdel-Fattah et al., 2015). A very simple antibody-antigen diagnostic based on specially treated paper to immobilize the antigen and a drinking straw as an incubation chamber has been developed for use in resource-limited setting (Chan et al., 2016) and might constitute an attractive basis for diagnostics based on peptides identified by ORFeome phage display such as ours.

Even though the two identified peptides demonstrated good diagnostic performance even higher accuracy would be desirable for clinical application. Combination of these peptides with proteins of immunodominant properties, such as Lip32 may result in better coverage of pathogenic serovars. However, adding recombinant proteins to the assay would obviate the clear advantages of using peptide-based assays. Thus, to preserve the “peptide only” aspect of a new diagnostic, it will be

important to assess the diagnostic value of the two dominant antigenic epitopes of lipL32 (Lottersberger et al., 2009) in peptide form in order to assess whether combining them with our peptides would be useful. In addition, future work should include additional screens to identify other immunoreactive peptides that could synergize diagnostically with SIR16-D1 and/or SIR16-H1 or be superior by themselves.

Since our goal was to identify peptides for the diagnosis of leptospirosis in the acute phase, we evaluated the peptides for reactivity against IgM only. All patient sera were obtained from patients who presented to the health care system after an average of three to five days of illness symptoms. IgM is the first agglutinating antibody to develop 5 to 14 days after exposure to infection and diagnostically meaningful IgG levels appear 1-3 weeks later (Picardeau et al., 2014) It would now be interesting to test whether these peptides are also reactive with IgG and could therefore be used for seroepidemiological surveys, in addition to diagnostics in the acute phase.

It came as a surprise that all five selected peptides were derived from the Malaysian genomic library, but none from the WHO reference genomic library. This is probably because the Malaysian strains had undergone fewer passages in culture than the WHO strains, thus preserving antigen profiles more closely resembling natural infection. This observation has important implications for future work, as it clearly suggests that early passage strains would constitute a better source for ORFeome phage display libraries than extensively subcultured reference strains.

Peptide SIR16-H1 corresponded to a predicted protein of unknown function. In contrast, peptide SIR16-D1 turned out to be a fragment of Glucose-1-phosphate cytidylyltransferase, also known as CDP-glucose pyrophosphorylase, the product of the *rfbF* gene (Koropatkin et al., 2004; Koropatkin et al., 2005; Thorson et al., 1994) This protein can be found in 10 leptospiral species and also in some other bacteria. Even though this is an intracellular protein, it is quite feasible that it does lead to a humoral immune response as it might become exposed to the immune system during lysis of leptospirae, during phagocytosis by antigen-presenting cells, or even by being secreted from live leptospirae in the sense of a “moonlighting protein” (Henderson et al., 2014).

8.0 SUMMARY

8.1 Whole Genome Analysis of *Leptospira spp.*

Leptospira interrogans is a very dynamic bacterium that can survive in multiple hosts and environments. It causes a wide range of disease spectrum which probably relate to its serovar and strains diversity. Both method of whole genome sequence typings i.e. cgMLST and SNP analysis of *L. interrogans* showed comparable topology placement and congruousness to that of MLST. However, SNP analysis displayed higher resolution typing with better intra-species discriminatory effect as polymorphisms can be discovered in both coding and noncoding regions of the genome. Genome plasticity in the genomes of *L. interrogans* are massive, which suggested that HGT is the most important measurement of adaptation to sustain its survival in reservoir host. These vicious cycles perpetuate to the diversity of the genome in various strains of *L. interrogans*. In this study, it could be shown that plasmids and insertion sequences are the main element that contributes to the genomic plasticity. Almost all genomes possess two or more than plasmids that differ from the formerly known ones by at least 100kB, showing that plasmids are common in *L. interrogans*, but the genomic diversity within *L. interrogans species* is much higher than thought before. Due to the advent in the genetics technology, few virulent factors have been identified. However, there are still lacking of knowledge on virulence factors that determine the variation in its virulence. In the comparison between the most and least severe strains in this study, there is an interesting findings on the transcription of pathogenesis function may explain strain Langkawi

superiority in virulence compared to that of strain 1530d. At logarithmic phase of in vitro culture, there was significant upregulation of genes associated with chemotaxis, which suggested higher driving force to induce infection and host entrance in strain Langkawi. These findings however best confirmed by study of *L. interrogans* invasion in mammalian cells. This study was performed only in six strains of *L. interrogans* originated from Malaysia. Although this study covers various aspect of the disease i.e clinical features and animal study; and pathogens i.e. genomic and trasncriptomic, the findings should be inferred carefully to other strains in *L. interrogans* species. The subject of leptospirosis pathogenesis and virulence are still areas with full of hypothesis. We suggest for a comprehensive *In vivo* transcriptomic study of *L. interrogans* with of strains with different virulence spectrum in order to increase our understanding on its diversity in virulence. Besides that, more studies of higher number of strains with clonal complexes are highly needed to understand the genomic evolution difference of *L. interrogans* intra-species genome beyond SNP.

8.2 Biomarker Identification of *Leptospira spp.*

This is the first study of seroreactive peptides identified by a phage display approach using a combination of endemic *Leptospira spp.* in Malaysia. The synthetic peptides SIR16-D1 and SIR16-H1 showed good potential for the discrimination of acute phase leptospirosis and healthy patients and can form the basis for the development of peptide-based diagnostics for use in resource-limited settings and hot climates countries.

This study was limited by the number of sera selected for ELISA and also by the prevalence of serovars in two endemic regions in Malaysia i.e. Kota Bharu and Kota Kinabalu. Besides, the control group was recruited from healthy donors from a non-endemic region. This is because individuals who had been exposed to leptospirosis in endemic region can produce antibodies from the memory pool, which would lead to background reactivity and false positive results. In fact, it was reported that healthy individuals in high endemic region have a 15% prevalence of positive anti-leptospira antibodies detected by MAT (Reis et al., 2008). Evaluation of the peptides with sera from patients infected with other known tropical infection diseases such as Dengue Fever, Chikugunya, typhod and etc. should be included, as it is important to

assess their practicality and specificity as a leptospirosis diagnostic assay in populations exposed to pathogens that may cause serological cross reactivity. A more comprehensive study involving sera from patients and healthy individuals from various endemic and non-endemic countries should be included in the future.

As there are various serovars causing leptospirosis worldwide, we suggest to apply the ORFeome phage display screen to genomically more diverse isolates and to human sera collections from various endemic regions in order to warrant a more universal selection and characterization of the antigen repertoire.

9.0 REFERENCES

Abby, S. S., Cury, J., Guglielmini, J., Néron, B., Touchon, M., & Rocha, E. P. C. (2016). Identification of protein secretion systems in bacterial genomes. Sci Rep, 6, 23080.doi:10.1038/srep23080. https://www.nature.com/articles/srep23080#supplementary-information

Abdel-Fattah, M., Charmy, R., & Tawfik, A. R. (2015). Immune Reactivity of Synthetic Peptides Originated from Hypervariable Region 1 of Hepatitis C Synthetic Consensuses with Egyptian Sera Infected with Hepatitis C Virus Type 4. *Inter virology*, 58(4), 232-241. doi:10.1159/000437385

Adler, B. (2014). Pathogenesis of leptospirosis: Cellular and molecular aspects. *Vet Microbiol*, 172(3), 353-358. doi:https://doi.org/10.1016/j.vetmic.2014.06.015

Adler, B. (2015). Vaccines against leptospirosis. In *Leptospira* and Leptospirosis (pp. 251-272): Springer.

Adler, B., & de la Peña Moctezuma, A. (2010). *Leptospira* and leptospirosis. *Vet Microbiol*, 140(3), 287-296. doi:https://doi.org/10.1016/j.vetmic.2009.03.012

Adler, B., Lo, M., Seemann, T., & Murray, G. L. (2011). Pathogenesis of leptospirosis: the influence of genomics. *Vet Microbiol*, 153(1-2), 73-81. doi:10.1016/j.vetmic.2011.02.055

Agampodi, S. B., Matthias, M. A., Moreno, A. C., & Vinetz, J. M. (2012). Utility of Quantitative Polymerase Chain Reaction in Leptospirosis Diagnosis: Association of Level of Leptospiremia and Clinical Manifestations in Sri Lanka. *Clinical Infectious Diseases*, 54(9), 1249-1255. doi:10.1093/cid/cis035

Agampodi, S. B., Moreno, A. C., Vinetz, J. M., & Matthias, M. A. (2013). Utility and limitations of direct multi-locus sequence typing on qPCR-positive blood to determine infecting *Leptospira* strain. *Am J Trop Med Hyg*, 88(1), 184-185. doi:10.4269/ajtmh.2012.12-0526

Agampodi, S. B., Peacock, S. J., Thevanesam, V., Nugegoda, D. B., Smythe, L., Thaipadungpanit, J., . . . Vinetz, J. M. (2011). Leptospirosis outbreak in Sri Lanka in 2008: lessons for assessing the global burden of disease. *Am J Trop Med Hyg*, 85(3), 471-478. doi:10.4269/ajtmh.2011.11-0276

Ahmed, A., Engelberts, M. F., Boer, K. R., Ahmed, N., & Hartskeerl, R. A. (2009). Development and validation of a real-time PCR for detection of pathogenic *Leptospira* species in clinical materials. *PLoS One*, 4(9), e7093. doi:10.1371/journal.pone.0007093

Ahmed, A., Thaipadungpanit, J., Boonsilp, S., Wuthiekanun, V., Nalam, K., Spratt, B. G., . . . Peacock, S. J. (2011). Comparison of two multilocus sequence based genotyping schemes *for Leptospira* species. *PLoS Negl Trop Dis*, 5(11), e1374. doi:10.1371/journal.pntd.0001374

Ahmed, N., Devi, S. M., Valverde Mde, L., Vijayachari, P., Machang'u, R. S., Ellis, W. A., & Hartskeerl, R. A. (2006). Multilocus sequence typing method for identification and genotypic classification of pathogenic *Leptospira* species. *Ann Clin Microbiol Antimicrob*, 5, 28. doi:10.1186/1476-0711-5-28

Ahmed, S. A., Sandai, D. A., Musa, S., Hoe, C. H., Riadzi, M., Lau, K. L., & Tang, T. H. (2012). Rapid diagnosis of leptospirosis by multiplex PCR. *Malays J Med Sci,* 19(3), 9-16.

Akira, S., Uematsu, S., & Takeuchi, O. (2006). Pathogen Recognition and Innate Immunity. Cell, 124(4), 783-801. doi:https://doi.org/10.1016/j.cell.2006.02.015

Akmatov, M. K., Krebs, S., Preusse, M., Gatzemeier, A., Frischmann, U., Schughart, K., & Pessler, F. (2011). E-mail-based symptomatic surveillance combined with self-collection of nasal swabs: a new tool for acute respiratory infection epidemiology. *Int J Infect Dis*, 15(11), e799-803. doi:10.1016/j.ijid.2011.07.005

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. J. N. a. r. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. 25(17), 3389-3402.

Amilasan, A. S., Ujiie, M., Suzuki, M., Salva, E., Belo, M. C., Koizumi, N., . . . Ariyoshi, K. (2012). Outbreak of leptospirosis after flood, the Philippines, 2009. *Emerg Infect Dis*, 18(1), 91-94. doi:10.3201/eid1801.101892

Amran, F. (2012). Leptospirosis in Malaysia. Paper presented at the 12th Western Pacific Chemotherapy and Infectious Diseases Conference, Singapore.

Andreescu, N. (1990). A new prepatory method of thermically inactivated *Leptospira* Patoc antigen for rapid slide agglutination used as serosurvey test for human leptospiroses. Archives roumaines de pathologie experimentales et de microbiologie, 49(3), 223-227.

Andre-Fontaine, G., Aviat, F., & Thorin, C. (2015). Waterborne Leptospirosis: Survival and Preservation of the Virulence of Pathogenic *Leptospira spp*. in Fresh Water. *Curr Microbiol*, 71(1), 136-142. doi:10.1007/s00284-015-0836-4

Appassakij, H., Silpapojakul, K., Wansit, R., & Woodtayakorn, J. (1995). Evaluation of the immunofluorescent antibody test for the diagnosis of human leptospirosis. Am J Trop Med Hyg, 52(4), 340-343.

Arimitsu, Y., Kmety, E., Ananyina, Y., Baranton, G., Ferguson, I. R., Smythe, L., & Terpstra, W. J. (1994). Evaluation of the one-point microcapsule agglutination test for diagnosis of leptospirosis. Bull World Health Organ, 72(3), 395-399.

Aronesty, E. (2011). ea-utils: Command-line tools for processing biological sequencing data. Durham,NC: Expression Analysis.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., . . . Gasteiger, E. J. N. a. r. (2012). ExPASy: SIB bioinformatics resource portal. 40(W1), W597-W603.

Athanazio, D. A., Silva, E. F., Santos, C. S., Rocha, G. M., Vannier-Santos, M. A., McBride, A. J., . . . Reis, M. G. J. A. t. (2008). Rattus norvegicus as a model for persistent renal colonization by pathogenic *Leptospira interrogans*. 105(2), 176-180.

Auch, A. F., von Jan, M., Klenk, H.-P., & Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Standards in genomic sciences, 2(1), 117-134. doi:10.4056/sigs.531120

Aviat, F., Slamti, L., Cerqueira, G. M., Lourdault, K., & Picardeau, M. (2010). Expanding the genetic toolbox for *Leptospira* species by generation of fluorescent bacteria. Appl Environ Microbiol, 76(24), 8135-8142. doi:10.1128/aem.02199-10

Bajani, M. D., Ashford, D. A., Bragg, S. L., Woods, C. W., Aye, T., Spiegel, R. A., . . . Weyant, R. S. (2003). Evaluation of four commercially available rapid serologic tests for diagnosis of leptospirosis. J Clin Microbiol, 41(2), 803-809.

Baril, C., Herrmann, J. L., Richaud, C., Margarita, D., & Girons, I. S. (1992). Scattering of the rRNA genes on the physical map of the circular chromosome of *Leptospira interrogans* serovar icterohaemorrhagiae. 174(23), 7566-7571. doi:10.1128/jb.174.23.7566-7571.1992 %J Journal of Bacteriology

Batzios, S. P., Zafeiriou, D. I., & Papakonstantinou, E. J. F. l. (2013). Extracellular matrix components: An intricate network of possible biomarkers for lysosomal storage disorders? , 587(8), 1258-1267.

Bauer, D. F. (1972). Constructing Confidence Sets Using Rank Statistics. *Journal of the American Statistical Association*, 67(339), 687-690. doi:10.1080/01621459.1972.10481279

Becker, M., Felsberger, A., Frenzel, A., Shattuck, W. M., Dyer, M., Kugler, J., . . . Hust, M. (2015). Application of M13 phage display for identifying immunogenic proteins from tick (Ixodes scapularis) saliva. *BMC Biotechnol,* 15, 43. doi:10.1186/s12896-015-0167-3

Bejo, S. K., Amran, F., Zakaria, Z., Daud, A., Sahani, M., & Evie, K. (2017). Manual for Laboratory Diagnosis of Leptospirosis : One Health Approach.

Beranova-Giorgianni, S. (2003). Proteome analysis by two-dimensional gel electrophoresis and mass spectrometry: strengths and limitations. TrAC Trends in Analytical Chemistry, 22(5), 273-281. doi:https://doi.org/10.1016/S0165-9936(03)00508-9

Bharti, A. R., Nally, J. E., Ricaldi, J. N., Matthias, M. A., Diaz, M. M., Lovett, M. A., . . . Vinetz, J. M. (2003). Leptospirosis: a zoonotic disease of global importance. *Lancet Infect Dis*, 3(12), 757-771.

BLAST, G. J. S. F. (1997). PSI-BLAST: a new generation of protein database search programs Altschul. 3389-3402.

Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. J. M. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. 151(8), 2551-2561.

Boonsilp, S., Thaipadungpanit, J., Amornchai, P., Wuthiekanun, V., Bailey, M. S., Holden, M. T. G., . . . Peacock, S. J. (2013). A Single Multilocus Sequence Typing (MLST) Scheme for Seven Pathogenic *Leptospira* Species. *PLOS Neglected Tropical Diseases*, 7(1), e1954. doi:10.1371/journal.pntd.0001954

Bourhy, P., Bremont, S., Zinini, F., Giry, C., & Picardeau, M. (2011). Comparison of real-time PCR assays for the detection of pathogenic *Leptospira spp.* in blood and identification of variations in target sequences. doi:10.1128/JCM.02452-10 Journal of Clinical Microbiology

Bourhy, P., Collet, L., Brisse, S., & Picardeau, M. (2014). *Leptospira* mayottensis sp. nov., a pathogenic species of the genus *Leptospira* isolated from humans. International journal of systematic and evolutionary microbiology, 64(Pt 12), 4061-4067. doi:10.1099/ijs.0.066597-0

Bourhy, P., Louvel, H., Saint Girons, I., & Picardeau, M. (2005). Random Insertional Mutagenesis of *Leptospira interrogans* the Agent of Leptospirosis, Using a mariner Transposon. 187(9), 3255-3258. doi:10.1128/JB.187.9.3255-3258.2005 %J Journal of Bacteriology

Bourhy, P., & Saint Girons, I. (2000). Localization of the *Leptospira interrogans* metF gene on the CII secondary chromosome. *FEMS Microbiol Lett*, 191(2), 259-263. doi:10.1111/j.1574-6968.2000.tb09349.x

Boursaux-Eude, C., Girons, I. S., & Zuerner, R. (1995). IS1500, an IS3-like element from *Leptospira interrogans*. 141(9), 2165-2173. doi:doi:10.1099/13500872-141-9-2165

Breitling, F., Dübel, S., Seehaus, T., Klewinghaus, I., & Little, M. (1991). A surface expression vector for antibody screening. *Gene*, 104(2), 147-153. doi:https://doi.org/10.1016/0378-1119(91)90244-6

Brenner, D. J., Kaufmann, A. F., Sulzer, K. R., Steigerwalt, A. G., Rogers, F. C., & Weyant, R. S. (1999). Further determination of DNA relatedness between serogroups and serovars in the family Leptospiraceae with a proposal for *Leptospira alexanderi* sp. nov. and four new *Leptospira* genomospecies. 49(2), 839-858. doi:doi:10.1099/00207713-49-2-839

Brüssow, H., Canchaya, C., Hardt, W.-D. J. M., & reviews, m. b. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. 68(3), 560-602.

Bryant, J., Chewapreecha, C., & Bentley, S. D. J. F. m. (2012). Developing insights into the mechanisms of evolution of bacterial pathogens from whole-genome sequences. 7(11), 1283-1296.

Bulach, D. M., Kalambaheti, T., de la Peña-Moctezuma, A., Adler, B. J. J. o. m. m., & biotechnology. (2000). Lipopolysaccharide biosynthesis in *Leptospira.* 2(4), 375-380.

Bulach, D. M., Zuerner, R. L., Wilson, P., Seemann, T., McGrath, A., Cullen, P. A., . . . Adler, B. (2006). Genome reduction in *Leptospira borgpetersenii* reflects limited transmission potential. *Proc Natl Acad Sci* U S A, 103(39), 14560-14565. doi:10.1073/pnas.0603979103

Cagliero, J., Villanueva, S. Y. A., Matsui, M. J. F. i. c., & microbiology, i. (2018). *Leptospira* pathophysiology: into the storm of cytokines. 8, 204.

Cameron, C. E. (2015). Leptospiral Structure, Physiology, and Metabolism. In B. Adler (Ed.), *Leptospira* and Leptospirosis (pp. 21-41). Berlin, Heidelberg: Springer Berlin Heidelberg.

Castiblanco-Valencia, M. M., Fraga, T. R., Silva, L. B., Monaris, D., Abreu, P. A., Strobel, S., . . . Barbosa, A. S. (2012). Leptospiral immunoglobulin-like proteins interact with human complement regulators factor H, FHL-1, FHR-1, and C4BP. *J Infect Dis*, 205(6), 995-1004. doi:10.1093/infdis/jir875

Cerqueira, G. M., McBride, A. J., Hartskeerl, R. A., Ahmed, N., Dellagostin, O. A., Eslabao, M. R., & Nascimento, A. L. (2010). Bioinformatics describes novel Loci for high resolution discrimination of *Leptospira* isolates. *PLoS One*, 5(10), e15335. doi:10.1371/journal.pone.0015335

Cerqueira, G. M., McBride, A. J., Picardeau, M., Ribeiro, S. G., Moreira, A. N., Morel, V., . . . Dellagostin, O. A. (2009). Distribution of the leptospiral immunoglobulin-like (lig) genes in pathogenic *Leptospira* species and application of ligB to typing leptospiral isolates. *J Med Microbiol*, 58(Pt 9), 1173-1181. doi:10.1099/jmm.0.009175-0

Chaemchuen, S., Rungpragayphan, S., Poovorawan, Y., & Patarakul, K. (2011). Identification of candidate host proteins that interact with LipL32, the major outer membrane

protein of pathogenic *Leptospira*, by random phage display peptide library. *Vet Microbiol*, 153(1), 178-185. doi:https://doi.org/10.1016/j.vetmic.2011.04.030

Chakraborty, A., Miyahara, S., Villanueva, S. Y., Gloriani, N. G., Yoshida, S.-i. J. A. a., & chemotherapy. (2010). In vitro sensitivity and resistance of 46 *Leptospira* strains isolated from rats in the Philippines to 14 antimicrobial agents. 54(12), 5403-5405.

Chan, S. K., & Lim, T. S. (2016). A straw-housed paper-based colorimetric antibody–antigen sensor. *Analytical Methods,* 8(6), 1431-1436. doi:10.1039/C5AY01828E

Chang, R. S., Smith, D. J. W., McComb, D. E., Sharp, C. F., & Tonge, J. I. (1957). The Use of Erythrocyte Sensitizing Substance in the Diagnosis of Leptospiroses. 6(1), 101-107. doi:doi:https://doi.org/10.4269/ajtmh.1957.6.101

Chapon-Hervé, V., Akrim, M., Latifi, A., Williams, P., Lazdunski, A., & Bally, M. (1997). Regulation of the xcp secretion pathway by multiple quorum-sensing modulons in *Pseudomonas aeruginosa*. 24(6), 1169-1178. doi:doi:10.1046/j.1365-2958.1997.4271794.x

Charon, N., Greenberg, E., Koopman, M., & Limberger, R. J. R. i. m. (1992). Spirochete chemotaxis, motility, and the structure of the spirochetal periplasmic flagella. 143(6), 597-603.

Chassin, C., Picardeau, M., Goujon, J.-M., Bourhy, P., Quellard, N., Darche, S., . . . Lacroix-Lamandé, S. J. T. J. o. I. (2009). TLR4-and TLR2-mediated B cell responses control the clearance of the bacterial pathogen, *Leptospira interrogans*. *jimmunol*. 0900506.

Chen, Z.-Y., He, C.-Y., & Kay, D. M. A. (2005). Improved Production and Purification of Minicircle DNA Vector Free of Plasmid Bacterial Sequences and Capable of Persistent Transgene Expression In Vivo. 16(1), 126-131. doi:10.1089/hum.2005.16.126

Chiani, Y., Jacob, P., Varni, V., Landolt, N., Schmeling, M. F., Pujato, N., . . . Evolution. (2016). Isolation and clinical sample typing of human leptospirosis cases in Argentina. 37, 245-251.

Ching ATC, Fávaro RD, Lima SS, Chaves AdAM, de Lima MA, Nader HB, et al (2012). *Lepstospira interrogans* shotgun phage display identified LigB as a heparin-binding protein. Biochem Biophys Res Commun. 427(4):774–9. https://doi.org/10.1016/j.bbrc.2012.09.137. pmid:23044419

Chirathaworn, C., Supputtamongkol, Y., Lertmaharit, S., & Poovorawan, Y. (2016). Cytokine levels as biomarkers for leptospirosis patients. Cytokine, 85, 80-82. doi:https://doi.org/10.1016/j.cyto.2016.06.007

Chousterman, B. G., Swirski, F. K., & Weber, G. F. J. S. i. I. (2017). Cytokine storm and sepsis disease pathogenesis. 39(5), 517-528. doi:10.1007/s00281-017-0639-8

Choy, H. A. (2012). Multiple activities of LigB potentiate virulence of *Leptospira interrogans*: inhibition of alternative and classical pathways of complement. *PLoS One,* 7(7), e41566. doi:10.1371/journal.pone.0041566

Cieslewicz, M. J., Kasper, D. L., Wang, Y., & Wessels, M. R. J. J. o. B. C. (2001). Functional Analysis in Type Ia Group B Streptococcus of a Cluster of Genes Involved in Extracellular Polysaccharide Production by Diverse Species of Streptococci. 276(1), 139-146.

Connor, D. O., Zantow, J., Hust, M., Bier, F. F., & von Nickisch-Rosenegk, M. (2016). Identification of Novel Immunogenic Proteins of *Neisseria gonorrhoeae* by Phage Display. *PLoS One*, 11(2), e0148986. doi:10.1371/journal.pone.0148986

Corney, B. G., Colley, J., Djordjevic, S. P., Whittington, R., & Graham, G. C. (1993). Rapid identification of some *Leptospira* isolates from cattle by random amplified polymorphic DNA fingerprinting. 31(11), 2927-2932.

Corwin, A., Ryan, A., Bloys, W., Thomas, R., Deniega, B., & Watts, D. (1990). A waterborne outbreak of leptospirosis among United States military personnel in Okinawa, Japan. *Int J Epidemiol*, 19(3), 743-748.

Costa, F., Hagan, J. E., Calcagno, J., Kane, M., Torgerson, P., Martinez-Silveira, M. S., . . . Ko, A. I. (2015). Global Morbidity and Mortality of Leptospirosis: A Systematic Review. *PLOS Neglected Tropical Diseases*, 9(9), e0003898. doi:10.1371/journal.pntd.0003898

Costa, T. R. D., Felisberto-Rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M., & Waksman, G. (2015). Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology*, 13, 343. doi:10.1038/nrmicro3456

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., . . . Gautheret, D. J. N. a. r. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins.

Criss, A. K., Katz, B. Z., & Seifert, H. S. J. C. m. (2009). Resistance of *Neisseria gonorrhoeae* to non- oxidative killing by adherent human polymorphonuclear leucocytes. 11(7), 1074-1087.

Croda, J., Figueira, C. P., Wunder, E. A., Santos, C. S., Reis, M. G., Ko, A. I., & Picardeau, M. (2008). Targeted Mutagenesis in Pathogenic *Leptospira* Species: Disruption of the LigB Gene Does Not Affect Virulence in Animal Models of Leptospirosis. 76(12), 5826-5833. doi:10.1128/IAI.00989-08 %*J Infection and Immunity*

Cumberland, P., Everard, C. O., & Levett, P. N. (1999). Assessment of the efficacy of an IgM-elisa and microscopic agglutination test (MAT) in the diagnosis of acute leptospirosis. *Am J Trop Med Hyg*, 61(5), 731-734.

Cunnion, K. M., Lee, J. C., & Frank, M. M. (2001). Capsule production and growth phase influence binding of complement to *Staphylococcus aureus*. Infect Immun, 69(11), 6796-6803. doi:10.1128/iai.69.11.6796-6803.2001

Cunnion, K. M., Zhang, H. M., & Frank, M. M. (2003). Availability of complement bound to *Staphylococcus aureus* to interact with membrane complement receptors influences efficiency of phagocytosis. Infect Immun, 71(2), 656-662.

da Rocha Medeiros, F., Spichler, A., & Athanazio, D. A. J. A. t. (2010). Leptospirosis-associated disturbances of blood vessels, lungs and hemostasis. 115(1-2), 155-162.

Dama, E., Cornelie, S., Camara, M., Somda, M. B., Poinsignon, A., Ilboudo, H., . . . Bucheton, B. (2013). In Silico Identification of a Candidate Synthetic Peptide (Tsgf118–43) to Monitor Human Exposure to Tsetse Flies in West Africa. *PLOS Neglected Tropical Diseases*, 7(9), e2455. doi:10.1371/journal.pntd.0002455

Danckert, L., Hoppe, S., Bier, F. F., & von Nickisch-Rosenegk, M. J. M. A. (2014). Rapid identification of novel antigens of *Salmonella* Enteritidis by microarray-based immunoscreening. 181(13), 1707-1714. doi:10.1007/s00604-014-1197-6

D'Angelo, S., Velappan, N., Mignone, F., Santoro, C., Sblattero, D., Kiss, C., & Bradbury, A. R. J. B. G. (2011). Filtering "genic" open reading frames from genomic DNA samples for advanced annotation. 12(1), S5. doi:10.1186/1471-2164-12-s1-s5

Darling, A. C., Mau, B., Blattner, F. R., & Perna, N. T. J. G. r. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. 14(7), 1394-1403.

Darmon, E., Leach, D. R. J. M., & Reviews, M. B. (2014). Bacterial genome instability. 78(1), 1-39.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. J. B. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. 837-845.

DelVecchio, V. G., Connolly, J. P., Alefantis, T. G., Walz, A., Quan, M. A., Patra, G., . . . Mujer, C. V. (2006). Proteomic Profiling and Identification of Immunodominant Spore Antigens of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*. 72(9), 6355-6363. doi:10.1128/AEM.00455-06 %J *Applied and Environmental Microbiology*

Dikken, H., & Kmety, E. (1978). Chapter VIII Serological Typing Methods of Leptospires. In T. Bergan & J. R. Norris (Eds.), Methods in Microbiology (Vol. 11, pp. 259-307): Academic Press.

Dinges, M. M., Orwin, P. M., & Schlievert, P. M. (2000). Exotoxins of Staphylococcus aureus. Clin Microbiol Rev, 13(1), 16-34, table of contents.

Doungchawee, G., Kositanont, U., Niwetpathomwat, A., Inwisai, T., Sagarasaeranee, P., & Haake, D. A. (2008). Early diagnosis of leptospirosis by immunoglobulin M immunoblot testing. *Clin Vaccine Immunol*, 15(3), 492-498. doi:10.1128/cvi.00152-07

Dubois, M. E., Hammarlund, E., & Slifka, M. K. (2012). Optimization of Peptide-Based ELISA for Serological Diagnostics: A Retrospective Study of Human Monkeypox Infection. 12(5), 400-409. doi:10.1089/vbz.2011.0779

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, *32*(5), 1792-1797.

Eshghi, A., Becam, J., Lambert, A., Sismeiro, O., Dillies, M.-A., Jagla, B., . . . immunity. (2014). A putative regulatory genetic locus modulates virulence in the pathogen *Leptospira interrogans*. IAI. 01803-01814.

Eshghi, A., Lourdault, K., Murray, G. L., Bartpho, T., Sermswan, R. W., Picardeau, M., . . . immunity. (2012). *Leptospira interrogans* catalase is required for resistance to H_2O_2 and for virulence. IAI. 00466-00412.

Evangelista, K., Franco, R., Schwab, A., & Coburn, J. J. P. n. t. d. (2014). *Leptospira interrogans* binds to cadherins. 8(1), e2672.

Faine, S. (1957). Virulence in *Leptospira*. I. Reactions of guinea-pigs to experimental infection with *Leptospira icterohaemorrhagiae*. British journal of experimental pathology, 38(1), 1-7.

Faine, S. (Ed.) (1982). Guidelines for the control of leptospirosis (Vol. 67).

Faine, S., Adler, B., Bolin, C., Perolat, P. . (1999). *Leptospira* and Leptospirosis (Vol. 2nd ed.).

Faix, P. H., Burg, M. A., Gonzales, M., Ravey, E. P., Baird, A., & Larocca, D. (2004). Phage display of cDNA libraries: enrichment of cDNA expression using open reading frame selection. 36(6), 1018-1029. doi:10.2144/04366rr03

Feasey, N. A., Dougan, G., Kingsley, R. A., Heyderman, R. S., & Gordon, M. A. J. T. L. (2012). Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa. 379(9835), 2489-2499.

Fernandes, L. G., Siqueira, G. H., Teixeira, A. R., Silva, L. P., Figueredo, J. M., Cosate, M. R., . . . immunopathology. (2016). *Leptospira spp*.: novel insights into host–pathogen interactions. 176, 50-57.

Filloux, A., Michel, G., & Bally, M. (1998). GSP-dependent protein secretion in Gram-negative bacteria: the Xcp system of *Pseudomonas aeruginosa*. FEMS Microbiology Reviews, 22(3), 177-198. doi:10.1111/j.1574-6976.1998.tb00366.x

Fletcher, W. (1928). Recent Work on Leptospirosis, Tsutsugamusfai Disease, and Tropical Typhus in the Federated Malay States. Transactions of the Royal Society of Tropical Medicine and Hygiene, 21(4), 265-282 pp. doi:10.1016/S0035-9203(28)90019-X

Fukunaga, M., & Mifuchi, I. J. J. o. b. (1989). Unique organization of *Leptospira interrogans* rRNA genes. 171(11), 5763-5767.

Galloway, R. L., & Levett, P. N. (2008). Evaluation of a modified pulsed-field gel electrophoresis approach for the identification of *Leptospira* serovars. *Am J Trop Med Hyg*, 78(4), 628-632.

Galloway, R. L., & Levett, P. N. (2010). Application and validation of PFGE for serovar identification of *Leptospira* clinical isolates. *PLoS Negl Trop Dis*, 4(9). doi:10.1371/journal.pntd.0000824

Garba, B., Bahaman, A. R., Bejo, S. K., Zakaria, Z., Mutalib, A. R., & Bande, F. (2018). Major epidemiological factors associated with leptospirosis in Malaysia. *Acta Trop*, 178, 242-247. doi:10.1016/j.actatropica.2017.12.010

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. J. T. A. o. A. S. (2008). A weakly informative default prior distribution for logistic and other regression models. 2(4),1360-1383.

Gibson, T., Higgins, D., Thompson, J. J. P., & uk/pub/software/mac/clustalw/clustalx, d. a. a. f. f. e. a. (1994). ClustalX.

Goris, M., Leeflang, M., Boer, K., Goeijenbier, M., van Gorp, E., Wagenaar, J., . . . Parasitology. (2011). Establishment of valid laboratory case definition for human leptospirosis. 3(2).

Gouveia, E. L., Metcalfe, J., de Carvalho, A. L. F., Aires, T. S. F., Villasboas-Bisneto, J. C., Queirroz, A., . . . Ko, A. I. (2008). Leptospirosis-associated severe pulmonary hemorrhagic syndrome, Salvador, Brazil. *Emerging infectious diseases*, 14(3), 505-508. doi:10.3201/eid1403.071064

Gulati, S., & Gulati, A. J. L. I. O. O. o. I. C. S. (2012). Pulmonary manifestations of leptospirosis. 29(4), 347.

Gupta, A., Shrivastava, N., Grover, P., Singh, A., Mathur, K., Verma, V., . . . Chaudhary, V. K. (2013). A Novel Helper Phage Enabling Construction of Genome-Scale ORF-Enriched Phage Display Libraries. *PLoS One*, 8(9), e75212. doi:10.1371/journal.pone.0075212

Haake, D. A., Chao, G., Zuerner, R. L., Barnett, J. K., Barnett, D., Mazel, M., . . . Bolin, C. A. (2000). The leptospiral major outer membrane protein LipL32 is a lipoprotein expressed during mammalian infection. *Infect Immun*, 68(4), 2276-2285.

Haake, D. A., & Levett, P. N. (2015). Leptospirosis in humans. *Curr Top Microbiol Immunol*, 387, 65-97. doi:10.1007/978-3-662-45059-8_5

Haake, D. A., & Matsunaga, J. (2010). *Leptospira*: a spirochaete with a hybrid outer membrane. *Mol Microbiol*, 77(4), 805-814. doi:10.1111/j.1365-2958.2010.07262.x

Hagan, J. E., Moraga, P., Costa, F., Capian, N., Ribeiro, G. S., Wunder, E. A., Jr., . . . Ko, A. I. (2016). Spatiotemporal Determinants of Urban Leptospirosis Transmission: Four-Year Prospective Cohort Study of Slum Residents in Brazil. *PLOS Neglected Tropical Diseases*, 10(1), e0004275. doi:10.1371/journal.pntd.0004275

He, P., Sheng, Y.-Y., Shi, Y.-Z., Jiang, X.-G., Qin, J.-H., Zhang, Z.-M., . . . Guo, X.-K. J. B. G. (2007). Genetic diversity among major endemic strains of *Leptospira interrogans* in China. 8(1), 204. doi:10.1186/1471-2164-8-204

Helmerhorst, H., Van Tol, E., Tuinman, P., De Vries, P., Hartskeerl, R., Grobusch, M., & Hovius, J. J. N. J. M. (2012). Severe pulmonary manifestation of leptospirosis. 70(5), 215-221.

Henderson, B., & Martin, A. C. (2014). Protein moonlighting: a new factor in biology and medicine. In: Portland Press Limited.

Herrmann, J. L., Baril, C., Bellenger, E., Perolat, P., Baranton, G., & Saint Girons, I. (1991). Genome conservation in isolates of *Leptospira interrogans*. 173(23), 7582-7588. doi:10.1128/jb.173.23.7582-7588.1991 %J *Journal of Bacteriology*

Herrmann, J. L., Bellenger, E., Perolat, P., Baranton, G., & Saint Girons, I. (1992). Pulsed-field gel electrophoresis of NotI digests of leptospiral DNA: a new rapid method of serovar identification. *J Clin Microbiol*, 30(7), 1696-1702.

Herrmann-Storck, C., Saint Louis, M., Foucand, T., Lamaury, I., Deloumeaux, J., Baranton, G., . . . Salin, J. J. E. i. d. (2010). Severe leptospirosis in hospitalized patients, Guadeloupe. 16(2), 331.

Hookey, J. V., Bryden, J., & Gatehouse, L. (1993). The use of 16S rDNA sequence analysis to investigate the phylogeny of Leptospiraceae and related spirochaetes. *J Gen Microbiol*, 139(11), 2585-2590. doi:10.1099/00221287-139-11-2585

Hoppe, S., Bier, F. F., & Nickisch-Rosenegk, M. v. (2013). Rapid Identification of Novel Immunodominant Proteins and Characterization of a Specific Linear Epitope of Campylobacter jejuni. *PLoS One*, 8(5), e65837. doi:10.1371/journal.pone.0065837

Hosmer D, Lemeshow SJNYJW, Sons. Applied logistic regression. 1989. 1989.

Huerta-Cepas, J., Forslund, K., Szklarczyk, D., Jensen, L., von Mering, C., & Bork, P. (2016). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. bioRxiv, 76331.

Human Leptospirosis: guidance for diagnosis, surveillance and control. (2003). WHO CDS CSR EPS, 23.

Hust, M., Dubel, S., & Schirrmann, T. (2007). Selection of recombinant antibodies from antibody gene libraries. Methods Mol Biol, 408, 243-255. doi:10.1007/978-1-59745-547-3_14

Hust, M., Meysing, M., Schirrmann, T., Selke, M., Meens, J., Gerlach, G. F., & Dubel, S. (2006). Enrichment of open reading frames presented on bacteriophage M13 using hyperphage. *Biotechniques,* 41(3), 335-342. doi:10.2144/000112225

Isogai, E., Isogai, H., Kurebayashi, Y., & Ito, N. J. Z. f. B., Mikrobiologie und Hygiene. Series A: Medical Microbiology, *Infectious Diseases, Virology*, Parasitology. (1986). Biological activities of leptospiral lipopolysaccharide. 261(1), 53-64.

Janowicz, A., De Massis, F., Ancora, M., Cammà, C., Patavino, C., Battisti, A., . . . Garofolo, G. (2018). Core Genome Multi Locus Sequence Typing and Single Nucleotide Polymorphism Analyses in the Epidemiology of *Brucella melitensis* Infections. doi:10.1128/JCM.00517-18 %J *Journal of Clinical Microbiology*

Jerse, A. E., Sharma, N. D., Simms, A. N., Crow, E. T., Snyder, L. A., & Shafer, W. M. (2003). A Gonococcal Efflux Pump System Enhances Bacterial Survival in a Female Mouse Model of Genital Tract Infection. 71(10), 5576-5582. doi:10.1128/IAI.71.10.5576-5582.2003 %*J Infection and Immunity*

Johnson, S., Steiner, B., Cruce, D., Perkins, G., Arko, R. J. I., & immunity. (1993). Characterization of a catalase-deficient strain of *Neisseria gonorrhoeae*: evidence for the significance of catalase in the biology of *N. gonorrhoeae*. 61(4), 1232-1238.

Junemann, S., Sedlazeck, F. J., Prior, K., Albersmeier, A., John, U., Kalinowski, J., . . . Harmsen, D. (2013). Updating benchtop sequencing performance comparison. *Nat Biotechnol*, 31(4), 294-296. doi:10.1038/nbt.2522

Kassegne, K., Hu, W., Ojcius, D. M., Sun, D., Ge, Y., Zhao, J., . . . Yan, J. J. T. J. o. i. d. (2013). Identification of collagenase as a critical virulence factor for invasiveness and transmission of pathogenic *Leptospira species*. 209(7), 1105-1115.

Katz, A. R., Ansdell, V. E., Effler, P. V., Middleton, C. R., & Sasaki, D. M. (2001). Assessment of the clinical presentation and treatment of 353 cases of laboratory-confirmed leptospirosis in Hawaii, 1974-1998. *Clin Infect Dis*, 33(11), 1834-1841. doi:10.1086/324084

King, A. M., Pretre, G., Bartpho, T., Sermswan, R. W., Toma, C., Suzuki, T., . . . immunity. (2014). High-temperature protein G is an essential virulence factor of *Leptospira interrogans*. 82(3), 1123-1131.

Kmety, E., Galton, M. M., & Sulzer, C. R. (1970). Further standardization of the agglutinin-absorption test in the serology of leptospires. *Bull World Health Organ*, 42(5), 733-738.

Ko, A. I., Goarant, C., & Picardeau, M. (2009). *Leptospira*: the dawn of the molecular genetics era for an emerging zoonotic pathogen. *Nat Rev Microbiol*, 7(10), 736-747. doi:10.1038/nrmicro2208

Kobayashi, Y. J. J. o. *I., & Chemotherapy*. (2001). Clinical observation and treatment of leptospirosis. 7(2), 59-68.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., . . . Wilson, R. K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*, 22(3), 568-576. doi:10.1101/gr.129684.111

Koizumi, N., Nakajima, C., Harunari, T., Tanikawa, T., Tokiwa, T., Uchimura, E., . . . Suzuki, Y. (2012). A new loop-mediated isothermal amplification method for rapid, simple, and sensitive detection of *Leptospira spp*. in urine. doi:10.1128/JCM.00481-12 %J Journal of Clinical Microbiology

Koizumi, N., & Watanabe, H. (2003). Molecular cloning and characterization of a novel leptospiral lipoprotein with OmpA domain. FEMS *Microbiol Lett,* 226(2), 215-219. doi:10.1016/S0378-1097(03)00619-0

Koropatkin, N. M., Cleland, W. W., & Holden, H. M. J. J. o. B. C. (2005). Kinetic and structural analysis of alpha-D-Glucose-1-phosphate cytidylyltransferase from *Salmonella typhi*.

Koropatkin, N. M., & Holden, H. M. J. J. o. B. C. (2004). Molecular Structure of α-d-Glucose-1-phosphate Cytidylyltransferase from *Salmonella typhi.* 279(42), 44023-44029.

Kügler, J., Nieswandt, S., Gerlach, G. F., Meens, J., Schirrmann, T., Hust, M. J. A. M., & Biotechnology. (2008). Identification of immunogenic polypeptides from a *Mycoplasma hyopneumoniae* genome library by phage display. 80(3), 447. doi:10.1007/s00253-008-1576-1

Kügler, J., Zantow, J., Meyer, T., & Hust, M. (2013). Oligopeptide M13 Phage Display in Pathogen Research. 5(10), 2531.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), 1870-1874. doi:10.1093/molbev/msw054

Kumar, S., Tamura, K., & Nei, M. J. K., Molecular Evolutionary Genetics Analysis. (1993). MEGA: Molecular Evolutionary Genetics Analysis (Pennsylvania State Univ., University Park).

Kunnath-Velayudhan, S., Salamon, H., Wang, H.-Y., Davidow, A. L., Molina, D. M., Huynh, V. T., . . . Gennaro, M. L. (2010). Dynamic antibody responses to the *Mycobacterium tuberculosis* proteome. 107(33), 14703-14708. doi:10.1073/pnas.1009080107 %J Proceedings of the National Academy of Sciences

La Scola, B., Bui, L. T., Baranton, G., Khamis, A., & Raoult, D. (2006). Partial rpoB gene sequencing for identification of *Leptospira* species. *FEMS Microbiol Lett*, 263(2), 142-147. doi:10.1111/j.1574-6968.2006.00377.x

LaFrentz, B. R., LaPatra, S. E., Call, D. R., Wiens, G. D., & Cain, K. D. (2011). Identification of immunogenic proteins within distinct molecular mass fractions of *Flavobacterium psychrophilum*. 34(11), 823-830. doi:doi:10.1111/j.1365-2761.2011.01297.x

Lally, E. T., Kieba, I. R., Sato, A., Green, C. L., Rosenbloom, J., Korostoff, J., . . . Robinson, M. K. J. J. o. B. C. (1997). RTX toxins recognize a β2 integrin on the surface of human target cells. 272(48), 30463-30469.

Lambert, A., Picardeau, M., Haake, D. A., Sermswan, R. W., Srikram, A., Adler, B., . . . immunity. (2012). FlaA proteins in *Leptospira interrogans* are essential for motility and virulence, but not required for the formation of flagella sheath. IAI. 00131-00112.

Lambert, A., Takahashi, N., Charon, N., & Picardeau, M. (2012). Chemotactic behavior of pathogenic and non-pathogenic *Leptospira* species. doi:10.1128/AEM.02288-12 %J Applied and Environmental Microbiology

Lau, C. L., Skelly, C., Smythe, L. D., Craig, S. B., & Weinstein, P. (2012). Emergence of new leptospiral serovars in American Samoa - ascertainment or ecological change? *BMC Infect Dis*, 12, 19. doi:10.1186/1471-2334-12-19

Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P. F., & Prohaska, S. J. J. B. B. (2011). Proteinortho: Detection of (Co-)orthologs in large-scale analysis. 12(1), 124. doi:10.1186/1471-2105-12-124

Lessa-Aquino, C., Borges Rodrigues, C., Pablo, J., Sasaki, R., Jasinskas, A., Liang, L., . . . Felgner, P. L. (2013). Identification of seroreactive proteins of *Leptospira* interrogans serovar copenhageni using a high-density protein microarray approach. *PLoS Negl Trop Dis*, 7(10), e2499. doi:10.1371/journal.pntd.0002499

Levett, P. N. (2001). Leptospirosis. *Clin Microbiol Rev*, 14(2), 296-326. doi:10.1128/cmr.14.2.296-326.2001

Levett, P. N. (2015). Systematics of leptospiraceae. In *Leptospira* and Leptospirosis (pp. 11-20): Springer.

Levin, B. R., & Bergstrom, C. T. J. P. o. t. N. A. o. S. (2000). Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. 97(13), 6981-6985.

Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A., & Dewey, C. N. J. B. (2009). RNA-Seq gene expression estimation with read mapping uncertainty. 26(4), 493-500.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Li, S., Ojcius, D. M., Liao, S., Li, L., Xue, F., Dong, H., & Yan, J. J. I. i. (2010). Replication or death: distinct fates of pathogenic *Leptospira* strain Lai within macrophages of human or mouse origin. 16(2), 80-92.

Liao, S., Sun, A., Ojcius, D. M., Wu, S., Zhao, J., & Yan, J. J. B. m. (2009). Inactivation of the fliY gene encoding a flagellar motor switch protein attenuates mobility and virulence of *Leptospira interrogans* strain Lai. 9(1), 253.

Lim, V. K. (2011). Leptospirosis: a re-emerging infection. *Malays J Pathol*, 33(1), 1-5.

Lima SS, Ching ATC, Fávaro RD, Da Silva JB, Oliveira MLS, Carvalho E, et al. (2013). Adhesin activity of *Leptospira interrogans* lipoprotein identified by in vivo and in vitro shotgun phage display. Biochem Biophys Res Commun. 431(2):342–7. https://doi.org/10.1016/j.bbrc.2012.12.095. pmid:23291183

Lin, X., Chen, Y., Lu, Y., Yan, J., & Yan, J. (2009). Application of a loop-mediated isothermal amplification method for the detection of pathogenic *Leptospira*. *Diagn Microbiol Infect Dis*, 63(3), 237-242. doi:10.1016/j.diagmicrobio.2008.10.012

Lin, Y.-P., McDonough, S. P., Sharma, Y., & Chang, Y.-F. J. P. o. (2010). The terminal immunoglobulin-like repeats of LigA and LigB of *Leptospira* enhance their binding to gelatin binding domain of fibronectin and host cells. 5(6), e11301.

Lingappa, J., Kuffner, T., Tappero, J., Whitworth, W., Mize, A., Kaiser, R., & McNicholl, J. (2004). HLA-DQ6 and ingestion of contaminated water: possible gene-environment interaction in an outbreak of Leptospirosis. *Genes Immun*, 5(3), 197-202. doi:10.1038/sj.gene.6364058

Llanes, A., Restrepo, C. M., & Rajeev, S. (2016). Whole Genome Sequencing Allows Better Understanding of the Evolutionary History of *Leptospira interrogans* Serovar Hardjo. PLoS One, 11(7), e0159387. doi:10.1371/journal.pone.0159387

Lottersberger, J., Guerrero, S. A., Tonarelli, G. G., Frank, R., Tarabla, H., & Vanasco, N. B. (2009). Epitope mapping of pathogenic *Leptospira* LipL32. 49(5), 641-645. doi:doi:10.1111/j.1472-765X.2009.02723.x

Lourdault, K., Cerqueira, G., Wunder, E. A., & Picardeau, M. (2011). Inactivation of clpB in the pathogen *Leptospira interrogans* reduces virulence and resistance to stress conditions. doi:10.1128/IAI.05168-11 %J Infection and Immunity

Lourdault, K., Cerqueira, G., Wunder, E. A., Picardeau, M. J. I., & immunity. (2011). Inactivation of clpB in the pathogen *Leptospira interrogans* reduces virulence and resistance to stress conditions. IAI. 05168-05111.

Majed, Z., Bellenger, E., Postic, D., Pourcel, C., Baranton, G., & Picardeau, M. (2005). Identification of Variable-Number Tandem-Repeat Loci in *Leptospira interrogans* Sensu Stricto. 43(2), 539-545. doi:10.1128/JCM.43.2.539-545.2005 %J Journal of Clinical Microbiology

Marasinghe, M., Bandara, AWMKK, Weerasekera MM, Gunasekara, TDCP, Ranasinghe N, Hapugoda MD, Fernando, SSN. (2017). Predictors of Disease Severity of Leptospirosis on Admission - A Sri Lankan Study. Paper presented at the 10th International Leptospirosis Society Conference, Palmerston North, New Zealand.

Marín-León, I., Pérez-Lozano, M. J., de Villar-Conde, E., Dastis-Bendala, C., Vargas-Romero, J., & Pumarola-Suñé, T. (1997). Prospective evaluation of the macroagglutination slide test for *Leptospira*. Serodiagnosis and Immunotherapy in Infectious Disease, 8(3), 191-193. doi:https://doi.org/10.1016/S0888-0786(96)01078-5

Marques, M. B., Kasper, D. L., Pangburn, M. K., & Wessels, M. R. (1992). Prevention of C3 deposition by capsular polysaccharide is a virulence mechanism of type III group B streptococci. 60(10), 3986-3993.

Marraffini, L. A., & Sontheimer, E. J. J. N. R. G. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. 11(3), 181.

Matsunaga, J., Lo, M., Bulach, D. M., Zuerner, R. L., Adler, B., Haake, D. A. J. I., & immunity. (2007). Response of *Leptospira interrogans* to physiologic osmolarity: relevance in signaling the environment-to-host transition. 75(6), 2864-2874.

Matsunaga, J., Young, T. A., Barnett, J. K., Barnett, D., Bolin, C. A., & Haake, D. A. (2002). Novel 45-kilodalton leptospiral protein that is processed to a 31-kilodalton growth-phase-regulated peripheral membrane protein. *Infect Immun*, 70(1), 323-334.

McBride, A. J., Athanazio, D. A., Reis, M. G., & Ko, A. I. (2005). Leptospirosis. *Curr Opin Infect Dis*, 18(5), 376-386.

McBride, A. J. A., Santos, B. L., Queiroz, A., Santos, A. C., Hartskeerl, R. A., Reis, M. G., & Ko, A. I. (2007). Evaluation of Four Whole-Cell *Leptospira* Based Serological Tests for Diagnosis of Urban Leptospirosis. 14(9), 1245-1248. doi:10.1128/CVI.00217-07 %J Clinical and Vaccine Immunology

McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C. A., . . . Tjaden, B. (2013). Computational analysis of bacterial RNA-Seq data. Nucleic Acids Res, 41(14), e140. doi:10.1093/nar/gkt444

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., & Göker, M. J. B. B. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. 14(1), 60. doi:10.1186/1471-2105-14-60

Mérien, F., Amouriaux, P., Perolat, P., Baranton, G., & Saint Girons, I. (1992). Polymerase chain reaction for detection of *Leptospira spp*. in clinical samples. 30(9), 2219-2224.

Merien, F., Portnoi, D., Bourhy, P., Charavay, F., Berlioz-Arthaud, A., & Baranton, G. (2005). A rapid and quantitative method for the detection of *Leptospira* species in human leptospirosis. FEMS Microbiol Lett, 249(1), 139-147. doi:10.1016/j.femsle.2005.06.011

Meyer, T., Schirrmann, T., Frenzel, A., Miethe, S., Stratmann-Selke, J., Gerlach, G. F., . . . Hust, M. (2012). Identification of immunogenic proteins and generation of antibodies against *Salmonella Typhimurium* using phage display. BMC Biotechnol, 12, 29. doi:10.1186/1472-6750-12-29

Mikulski, M., Boisier, P., Lacassin, F., Soupé-Gilbert, M.-E., Mauron, C., Bruyere-Ostells, L., . . . Diseases, I. (2015). Severity markers in severe leptospirosis: a cohort study. 34(4), 687-695. doi:10.1007/s10096-014-2275-8

Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J., & Almendros, C. J. M. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. 155(3), 733-740.

Mojica, F. J., García-Martínez, J., & Soria, E. J. J. o. m. e. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. 60(2), 174-182.

Moll, H. J. C. M. (2003). Dendritic cells and host resistance to infection. 5(8), 493-500.

Moncrief, M. B. C., & Maguire, M. E. J. J. J. o. B. I. C. (1999). Magnesium transport in prokaryotes. 4(5), 523-527. doi:10.1007/s007750050374

Moon, K. H., Zhao, X., Manne, A., Wang, J., Yu, Z., Liu, J., & Motaleb, M. A. (2016). Spirochetes flagellar collar protein FlbB has astounding effects in orientation of periplasmic

flagella, bacterial shape, motility, and assembly of motors in *Borrelia burgdorferi*. 102(2), 336-348. doi:doi:10.1111/mmi.13463

Morey, R. E., Galloway, R. L., Bragg, S. L., Steigerwalt, A. G., Mayer, L. W., & Levett, P. N. (2006). Species-specific identification of Leptospiraceae by 16S rRNA gene sequencing. J Clin Microbiol, 44(10), 3510-3516. doi:10.1128/jcm.00670-06

Mounier, J., Bahrani, F. K., Sansonetti, P. J. J. I., & immunity. (1997). Secretion of Shigella flexneri Ipa invasins on contact with epithelial cells and subsequent entry of the bacterium into cells are growth stage dependent. 65(2), 774-782.

Muench, D. F., Kuch, D. J., Wu, H., Begum, A. A., Veit, S. J., Pelletier, M.-E., . . . Jerse, A. E. J. T. J. o. i. d. (2009). Hydrogen peroxide–producing lactobacilli Inhibit gonococci in vitro but not during experimental genital tract infection. 199(9), 1369-1378.

Murray, G. L. (2015). The molecular basis of leptospiral pathogenesis. In *Leptospira* and Leptospirosis (pp. 139-185): Springer.

Murray, G. L., Ellis, K. M., Lo, M., & Adler, B. (2008*). Leptospira interrogans* requires a functional heme oxygenase to scavenge iron from hemoglobin. Microbes and Infection, 10(7), 791-797. doi:https://doi.org/10.1016/j.micinf.2008.04.010

Murray, G. L., Morel, V., Cerqueira, G. M., Croda, J., Srikram, A., Henry, R., . . . Picardeau, M. (2009). Genome-Wide Transposon Mutagenesis in Pathogenic *Leptospira* Species. 77(2), 810-816. doi:10.1128/IAI.01293-08 %J Infection and Immunity

Murray, G. L., Srikram, A., Henry, R., Hartskeerl, R. A., Sermswan, R. W., & Adler, B. (2010). Mutations affecting *Leptospira interrogans* lipopolysaccharide attenuate virulence. 78(3), 701-709. doi:doi:10.1111/j.1365-2958.2010.07360.x

Murray, G. L., Srikram, A., Henry, R., Puapairoj, A., Sermswan, R. W., & Adler, B. (2009). *Leptospira interrogans* requires heme oxygenase for disease pathogenesis. Microbes and Infection, 11(2), 311-314. doi:https://doi.org/10.1016/j.micinf.2008.11.014

Murray, G. L., Srikram, A., Hoke, D. E., Wunder, E. A., Henry, R., Lo, M., . . . Adler, B. (2009). Major Surface Protein LipL32 Is Not Required for Either Acute or Chronic Infection with *Leptospira interrogans*. 77(3), 952-958. doi:10.1128/IAI.01370-08 %J Infection and Immunity

Murray, G. L., Srikram, A., Hoke, D. E., Wunder, E. A., Henry, R., Lo, M., . . . Adler, B. (2009). Major Surface Protein LipL32 Is Not Required for Either Acute or Chronic Infection with *Leptospira interrogans*. 77(3), 952-958. doi:10.1128/IAI.01370-08 %J Infection and Immunity

Mwachui, M. A., Crump, L., Hartskeerl, R., Zinsstag, J., & Hattendorf, J. (2015). Environmental and Behavioural Determinants of Leptospirosis Transmission: A Systematic Review. PLOS Neglected Tropical Diseases, 9(9), e0003843. doi:10.1371/journal.pntd.0003843

Nair, S., Milohanic, E., Berche, P. J. I., & immunity. (2000). ClpC ATPase is required for cell adhesion and invasion of *Listeria monocytogenes*. 68(12), 7061-7068.

Nally, J. E., Chantranuwat, C., Wu, X.-Y., Fishbein, M. C., Pereira, M. M., Da Silva, J. J. P., . . . Lovett, M. A. (2004). Alveolar septal deposition of immunoglobulin and complement

parallels pulmonary hemorrhage in a guinea pig model of severe pulmonary leptospirosis. The American journal of pathology, 164(3), 1115-1127. doi:10.1016/S0002-9440(10)63198-7

Nally, J. E., Chow, E., Fishbein, M. C., Blanco, D. R., Lovett, M. A. J. I., & immunity. (2005). Changes in lipopolysaccharide O antigen distinguish acute versus chronic *Leptospira interrogans* infections. 73(6), 3251-3260.

Narayanavari, S. A., Sritharan, M., Haake, D. A., & Matsunaga, J. J. M. (2012). Multiple leptospiral sphingomyelinases (or are there?). 158(5), 1137-1146.

Nascimento, A., Ko, A. I., Martins, E., Monteiro-Vitorello, C., Ho, P., Haake, D., . . . Oliveira, M. J. J. o. b. (2004). Comparative genomics of two *Leptospira interrogans* serovars reveals novel insights into physiology and pathogenesis. 186(7), 2164-2172.

Nascimento, A. L. T. O., Verjovski-Almeida, S., Van Sluys, M. A., Monteiro-Vitorello, C. B., Camargo, L. E. A., Digiampietri, L. A., . . . Martins, E. A. L. (2004). Genome features of *Leptospira interrogans* serovar Copenhageni. Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas, 37(4), 459-477.

Naseem, S., Meens, J., Jores, J., Heller, M., Dubel, S., Hust, M., & Gerlach, G. F. (2010). Phage display-based identification and potential diagnostic application of novel antigens from *Mycoplasma mycoides* subsp. mycoides small colony type. Vet Microbiol, 142(3-4), 285-292. doi:10.1016/j.vetmic.2009.09.071

National Antibiotic Guideline (Vol. 2nd edition). (2014): Ministry of Health, Malaysia.

Panaphut, T., Domrongkitchaiporn, S., Vibhagool, A., Thinkamrop, B., & Susaengrat, W. J. C. I. D. (2003). Ceftriaxone compared with sodium penicillin G for treatment of severe leptospirosis. 36(12), 1507-1513.

Papa, A., & Kotrotsiou, T. (2015). Cytokines in human leptospirosis. Transactions of the Royal Society of Tropical Medicine and Hygiene, 109(12), 749-754. doi:10.1093/trstmh/trv095

Park, S.-K., Lee, S.-H., Rhee, Y.-K., Kang, S.-K., Kim, K.-J., Kim, M.-C., . . . hygiene. (1989). Leptospirosis in Chonbuk Province of Korea in 1987: a study of 93 patients. 41(3), 345-351.

Paster, B. J., Dewhirst, F. E., Weisburg, W. G., Tordoff, L. A., Fraser, G. J., Hespell, R. B., . . . Woese, C. R. (1991). Phylogenetic analysis of the spirochetes. *J Bacteriol*, 173(19), 6101-6109.

Perez, J., Brescia, F., Becam, J., Mauron, C., & Goarant, C. (2011). Rodent abundance dynamics and leptospirosis carriage in an area of hyper-endemicity in New Caledonia. PLoS Negl Trop Dis, 5(10), e1361. doi:10.1371/journal.pntd.0001361

Perez, J., & Goarant, C. J. B. M. (2010). Rapid *Leptospira* identification by direct sequencing of the diagnostic PCR products in New Caledonia. 10(1), 325. doi:10.1186/1471-2180-10-325

Perolat, P., Chappel, R. J., Adler, B., Baranton, G., Bulach, D. M., Billinghurst, M. L., . . . Serrano, M. S. (1998). *Leptospira fainei* sp. nov., isolated from pigs in Australia. 48(3), 851-858. doi:doi:10.1099/00207713-48-3-851

Perolat, P., Merien, F., Ellis, W. A., & Baranton, G. (1994). Characterization of *Leptospira* isolates from serovar hardjo by ribotyping, arbitrarily primed PCR, and mapped restriction site polymorphisms. 32(8), 1949-1957.

Picardeau, M. (2017). Virulence of the zoonotic agent of leptospirosis: still terra incognita? Nat Rev Microbiol, 15(5), 297-307. doi:10.1038/nrmicro.2017.5

Picardeau, M., Bertherat, E., Jancloes, M., Skouloudis, A. N., Durski, K., & Hartskeerl, R. A. (2014). Rapid tests for diagnosis of leptospirosis: Current tools and emerging technologies. Diagnostic Microbiology and Infectious Disease, 78(1), 1-8. doi:https://doi.org/10.1016/j.diagmicrobio.2013.09.012

Picardeau, M., Brenot, A., & Saint Girons, I. (2001). First evidence for gene replacement in *Leptospira spp.* Inactivation of *L. biflexa* flaB results in non-motile mutants deficient in endoflagella. 40(1), 189-199. doi:doi:10.1046/j.1365-2958.2001.02374.x

Picardeau, M., Bulach, D. M., Bouchier, C., Zuerner, R. L., Zidane, N., Wilson, P. J., . . . Adler, B. (2008). Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of *Leptospira* and the pathogenesis of leptospirosis. PLoS One, 3(2), e1607. doi:10.1371/journal.pone.0001607

Pightling, A. W., Petronella, N., & Pagotto, F. J. B. r. n. (2015). Choice of reference-guided sequence assembler and SNP caller for analysis of *Listeria monocytogenes* short-read sequence data greatly influences rates of error. 8(1), 748.

Plank, R., Dean, D. J. M., & infection. (2000). Overview of the epidemiology, microbiology, and pathogenesis of *Leptospira spp.* in humans. 2(10), 1265-1276.

Pourcel, C., Salvignol, G., & Vergnaud, G. J. M. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. 151(3), 653-663.

Que-Gewirth, N. L., Ribeiro, A. A., Kalb, S. R., Cotter, R. J., Bulach, D. M., Adler, B., . . . Raetz, C. R. J. J. o. B. C. (2004). A methylated phosphate group and four amide-linked acyl chains in *Leptospira interrogans* lipid A The membrane anchor of an unusual lipopolysaccharide that activates TLR2. 279(24), 25420-25429.

Rabilloud, T., Chevallet, M., Luche, S., & Lelong, C. (2010). Two-dimensional gel electrophoresis in proteomics: Past, present and future. Journal of Proteomics, 73(11), 2064-2077. doi:https://doi.org/10.1016/j.jprot.2010.05.016

Ralph, D., McClelland, M., Welsh, J., Baranton, G., & Perolat, P. (1993). *Leptospira species* categorized by arbitrarily primed polymerase chain reaction (PCR) and by mapped restriction polymorphisms in PCR-amplified rRNA genes. 175(4), 973-981. doi:10.1128/jb.175.4.973-981.1993 %J Journal of Bacteriology

Ramadass, P., Jarvis, B. D. W., Corner, R. J., Penny, D., & Marshall, R. B. (1992). Genetic Characterization of Pathogenic *Leptospira Species* by DNA Hybridization. 42(2), 215-219. doi:doi:10.1099/00207713-42-2-215

Re, F., & Strominger, J. L. J. J. o. B. C. (2001). Toll-like receptor 2 (TLR2) and TLR4 differentially activate human dendritic cells. 276(40), 37692-37699.

Reis, E. A. G., Hagan, J. E., Ribeiro, G. S., Teixeira-Carvalho, A., Martins-Filho, O. A., Montgomery, R. R., . . . Reis, M. G. (2013). Cytokine Response Signatures in Disease

Progression and Development of Severe Clinical Outcomes for Leptospirosis. PLOS Neglected Tropical Diseases, 7(9), e2457. doi:10.1371/journal.pntd.0002457

Reis, R. B., Ribeiro, G. S., Felzemburgh, R. D. M., Santana, F. S., Mohr, S., Melendez, A. X. T. O., . . . Ko, A. I. (2008). Impact of Environment and Social Gradient on *Leptospira* Infection in Urban Slums. PLOS Neglected Tropical Diseases, 2(4), e228. doi:10.1371/journal.pntd.0000228

Ren, S.-X., Fu, G., Jiang, X.-G., Zeng, R., Miao, Y.-G., Xu, H., . . . Zhao, G.-P. (2003). Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. Nature, 422, 888. doi:10.1038/nature01597https://www.nature.com/articles/nature01597#supplementary-information

Rettinger, A., Krupka, I., Grunwald, K., Dyachenko, V., Fingerle, V., Konrad, R., . . . Huber, I. (2012). *Leptospira spp.* strain identification by MALDI TOF MS is an equivalent tool to 16S rRNA gene sequencing and multi locus sequence typing (MLST). BMC Microbiol, 12, 185. doi:10.1186/1471-2180-12-185

Ricaldi, J. N., Fouts, D. E., Selengut, J. D., Harkins, D. M., Patra, K. P., Moreno, A., . . . Matthias, M. A. (2012). Whole Genome Analysis of *Leptospira licerasiae* Provides Insight into Leptospiral Evolution and Pathogenicity. PLOS Neglected Tropical Diseases, 6(10), e1853. doi:10.1371/journal.pntd.0001853

Ricaldi, J. N., & Vinetz, J. M. J. C. I. D. R. (2006). Leptospirosis in the tropics and in travelers. 8(1), 51-58. doi:10.1007/s11908-006-0035-3

Richter, M., & Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A, 106(45), 19126-19131. doi:10.1073/pnas.0906412106

Ristow, P., Bourhy, P., Kerneis, S., Schmitt, C., Prevost, M.-C., Lilenbaum, W., & Picardeau, M. (2008). Biofilm formation by saprophytic and pathogenic leptospires. 154(5), 1309-1317. doi:doi:10.1099/mic.0.2007/014746-0

Robey, M., & Cianciotto, N. P. (2002). *Legionella pneumophila* feoAB Promotes Ferrous Iron Uptake and Intracellular Infection. 70(10), 5659-5669. doi:10.1128/IAI.70.10.5659-5669.2002 %J Infection and Immunity

Rolfe, M. D., Rice, C. J., Lucchini, S., Pin, C., Thompson, A., Cameron, A. D., . . . Baranyi, J. J. J. o. b. (2012). Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. 194(3), 686-701.

Romero, E. C., Blanco, R. M., & Galloway, R. L. (2009). Application of pulsed-field gel electrophoresis for the discrimination of leptospiral isolates in Brazil. Lett Appl Microbiol, 48(5), 623-627. doi:10.1111/j.1472-765X.2009.02580.x

Rondot, S., Koch, J., Breitling, F., & Dübel, S. (2001). A helper phage to improve single-chain antibody presentation in phage display. Nature Biotechnology, 19, 75. doi:10.1038/83567

Rouquette, C., De Chastellier, C., Nair, S., & Berche, P. J. M. m. (1998). The ClpC ATPase of *Listeria monocytogenes* is a general stress protein required for virulence and promoting early bacterial escape from the phagosome of macrophages. 27(6), 1235-1245.

Rouquette, C., Ripio, M. T., Pellegrini, E., Bolla, J. M., Tascon, R. I., Vázquez- Boland, J. A., & Berche, P. J. M. m. (1996). Identification of a ClpC ATPase required for stress tolerance and in vivo survival of *Listeria monocytogenes.* 21(5), 977-987.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., & Barrell, B. (2000). Artemis: sequence visualization and annotation. Bioinformatics, 16(10), 944-945. doi:10.1093/bioinformatics/16.10.944

Sagna, A. B., Gaayeb, L., Sarr, J. B., Senghor, S., Poinsignon, A., Boutouaba-Combe, S., . . . Riveau, G. J. M. J. (2013). *Plasmodium falciparum* infection during dry season: IgG responses to Anopheles gambiae salivary gSG6-P1 peptide as sensitive biomarker for malaria risk in Northern Senegal. 12(1), 301. doi:10.1186/1475-2875-12-301

Schiettekatte, O., Vincent, A. T., Malosse, C., Lechat, P., Chamot-Rooke, J., Veyrier, F. J., . . . Bourhy, P. J. S. r. (2018). Characterization of LE3 and LE4, the only lytic phages known to infect the spirochete *Leptospira*. 8.

Schmid, G. P., Steere, A. C., Kornblatt, A. N., Kaufmann, A. F., Moss, C. W., Johnson, R. C., . . . Brenner, D. J. (1986). Newly recognized *Leptospira species* ("*Leptospira inadai*" serovar lyme) isolated from human skin. J Clin Microbiol, 24(3), 484-486.

Schulte, W., Bernhagen, J., #xfc, rgen, & Bucala, R. (2013). Cytokines in Sepsis: Potent Immunoregulators and Potential Therapeutic Targets—An Updated View %J Mediators of Inflammation. 2013, 16. doi:10.1155/2013/165974

Schurch, A. C., Arredondo-Alonso, S., Willems, R. J. L., & Goering, R. V. (2018). Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. Clin Microbiol Infect, 24(4), 350-354. doi:10.1016/j.cmi.2017.12.016

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics, 30(14), 2068-2069. doi:10.1093/bioinformatics/btu153

Segura, E. R., Ganoza, C. A., Campos, K., Ricaldi, J. N., Torres, S., Silva, H., . . . Vinetz, J. M. (2005). Clinical spectrum of pulmonary involvement in leptospirosis in a region of endemicity, with quantification of leptospiral burden. Clin Infect Dis, 40(3), 343-351. doi:10.1086/427110

Seki, M., Sato, T., Arimitsu, Y., Matuhasi, T., & Kobayashi, S. (1987). One-point method for serological diagnosis of leptospirosis: a microcapsule agglutination test. Epidemiol Infect, 99(2), 399-405. doi:10.1017/S0950268800067881

Shafer, W. M., Qu, X.-D., Waring, A. J., & Lehrer, R. I. (1998). Modulation of *Neisseria gonorrhoeae* susceptibility to vertebrate antibacterial peptides due to a member of the resistance/nodulation/division efflux pump family. 95(4), 1829-1833. doi:10.1073/pnas.95.4.1829 %J Proceedings of the National Academy of Sciences

Signorini, M. L., Lottersberger, J., Tarabla, H. D., & Vanasco, N. B. (2013). Enzyme-linked immunosorbent assay to diagnose human leptospirosis: a meta-analysis of the published literature. Epidemiol Infect, 141(1), 22-32. doi:10.1017/s0950268812001951

Siguier, P., Pérochon, J., Lestrade, L., Mahillon, J., & Chandler, M. J. N. a. r. (2006). ISfinder: the reference centre for bacterial insertion sequences. 34(suppl_1), D32-D36.

Silva, H. R., Tanajura, G. M., Tavares-Neto, J., Gomes, M. d. L. C., Linhares, A. d. C., Vasconcelos, P. F., & Ko, A. I. J. R. d. S. B. d. M. T. (2002). Aseptic meningitis syndrome due to enterovirus and *Leptospira sp* in children of Salvador, Bahia. 35(2), 159-165.

Silva, M. V., Camargo, E. D., Batista, L., Vaz, A. J., Brandao, A. P., Nakamura, P. M., & Negrao, J. M. (1995). Behaviour of specific IgM, IgG and IgA class antibodies in human leptospirosis during the acute phase of the disease and during convalescence. J Trop Med Hyg, 98(4), 268-272.

Slack, A., Symonds, M., Dohnt, M., Harris, C., Brookes, D., & Smythe, L. (2007). Evaluation of a modified Taqman assay detecting pathogenic *Leptospira spp.* against culture and *Leptospira*-specific IgM enzyme-linked immunosorbent assay in a clinical environment. Diagn Microbiol Infect Dis, 57(4), 361-366. doi:10.1016/j.diagmicrobio.2006.10.004

Slack, A. T., Symonds, M. L., Dohnt, M. F., & Smythe, L. D. (2006). Identification of pathogenic *Leptospira species* by conventional or real-time PCR and sequencing of the DNA gyrase subunit B encoding gene. BMC Microbiol, 6, 95. doi:10.1186/1471-2180-6-95

Slamti, L., de Pedro, M. A., Guichet, E., & Picardeau, M. J. J. o. b. (2011). Deciphering morphological determinants of the helix-shaped *Leptospira.* JB. 05695-05611.

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science, 228(4705), 1315-1317.

Smits, H. L., Chee, H. D., Eapen, C. K., Kuriakose, M., Sugathan, S., Gasem, M. H., . . . Gussenhoven, G. C. (2001). Latex based, rapid and easy assay for human leptospirosis in a single test format. Trop Med Int Health, 6(2), 114-118.

Smythe, L. D., Smith, I. L., Smith, G. A., Dohnt, M. F., Symonds, M. L., Barnett, L. J., & McKay, D. B. J. B. I. D. (2002). A quantitative PCR (TaqMan) assay for pathogenic *Leptospira spp.* 2(1), 13. doi:10.1186/1471-2334-2-13

Soler-García, Á. A., & Jerse, A. E. J. M. p. (2004). A *Neisseria gonorrhoeae* catalase mutant is more sensitive to hydrogen peroxide and paraquat, an inducer of toxic oxygen radicals. 37(2), 55-63.

Soltes, G., Hust, M., Ng, K. K. Y., Bansal, A., Field, J., Stewart, D. I. H., . . . Wiersma, E. J. (2007). On the influence of vector design on antibody phage display. Journal of Biotechnology, 127(4), 626-637. doi:https://doi.org/10.1016/j.jbiotec.2006.08.015

Sonthayanon, P., Chierakul, W., Wuthiekanun, V., Thaipadungpanit, J., Kalambaheti, T., Boonsilp, S., . . . Peacock, S. J. (2011). Accuracy of Loop-Mediated Isothermal Amplification for Diagnosis of Human Leptospirosis in Thailand. 84(4), 614-620. doi:doi:https://doi.org/10.4269/ajtmh.2011.10-0473

Stern, E. J., Galloway, R., Shadomy, S. V., Wannemuehler, K., Atrubin, D., Blackmore, C., . . . Clark, T. A. (2010). Outbreak of leptospirosis among Adventure Race participants in Florida, 2005. Clin Infect Dis, 50(6), 843-849. doi:10.1086/650578

Stoddard, R. A., Gee, J. E., Wilkins, P. P., McCaustland, K., & Hoffmaster, A. R. (2009). Detection of pathogenic *Leptospira spp.* through TaqMan polymerase chain reaction targeting the LipL32 gene. Diagn Microbiol Infect Dis, 64(3), 247-255. doi:10.1016/j.diagmicrobio.2009.03.014

Stohl, E. A., & Seifert, H. S. J. J. o. b. (2006). *Neisseria gonorrhoeae* DNA recombination and repair enzymes protect against oxidative damage caused by hydrogen peroxide. 188(21), 7645-7651.

Stone, C. B., Johnson, D. L., Bulir, D. C., Gilchrist, J. D., & Mahony, J. B. (2008). Characterization of the Putative Type III Secretion ATPase CdsN (Cpn0707) of *Chlamydophila pneumoniae*. 190(20), 6580-6588. doi:10.1128/JB.00761-08 %J Journal of Bacteriology

Tahara, H., Takabe, K., Sasaki, Y., Kasuga, K., Kawamoto, A., Koizumi, N., & Nakamura, S. (2018). The mechanism of two-phase motility in the spirochete *Leptospira*: Swimming and crawling. 4(5). doi:10.1126/sciadv.aar7975 %J Science Advances

Takahashi, S., Aoyagi, Y., Adderson, E. E., Okuwaki, Y., Bohnsack, J. F. J. I., & immunity. (1999). Capsular sialic acid limits C5a production on type III group B streptococci. 67(4), 1866-1870.

Terpstra, W. J., Korver, H., van Leeuwen, J., Klatser, P. R., & Kolk, A. H. (1985). The classification of Sejroe group serovars of *Leptospira interrogans* with monoclonal antibodies. Zentralbl Bakteriol Mikrobiol Hyg A, 259(4), 498-506.

Thaipadungpanit, J., Chierakul, W., Wuthiekanun, V., Limmathurotsakul, D., Amornchai, P., Boonslip, S., . . . Peacock, S. J. (2011). Diagnostic accuracy of real-time PCR assays targeting 16S rRNA and lipL32 genes for human leptospirosis in Thailand: a case-control study. PLoS One, 6(1), e16236. doi:10.1371/journal.pone.0016236

Thakker, M., Park, J.-S., Carey, V., Lee, J. C. J. I., & immunity. (1998). *Staphylococcus aureus* serotype 5 capsular polysaccharide is antiphagocytic and enhances bacterial virulence in a murine bacteremia model. 66(11), 5183-5189.

Thorson, J. S., Kelly, T. M., & Liu, H.-W. J. J. o. b. (1994). Cloning, sequencing, and overexpression in *Escherichia coli* of the alpha-D-glucose-1-phosphate cytidylyltransferase gene isolated from Yersinia pseudotuberculosis. 176(7), 1840-1849.

Toma, C., Okura, N., Takayama, C., & Suzuki, T. J. C. m. (2011). Characteristic features of intracellular pathogenic *Leptospira* in infected murine macrophages. 13(11), 1783-1792.

Trevejo, R. T., Rigau-Pérez, J. G., Ashford, D. A., McClure, E. M., Jarquín-González, C., Amador, J. J., . . . Shieh, W.-J. J. T. J. o. i. d. (1998). Epidemic leptospirosis associated with pulmonary hemorrhage—Nicaragua, 1995. 178(5), 1457-1463.

Truccolo, J., Serais, O., Merien, F., & Perolat, P. (2001). Following the course of human leptospirosis: evidence of a critical threshold for the vital prognosis using a quantitative PCR assay. FEMS Microbiol Lett, 204(2), 317-321.

Trueba, G., Zapata, S., Madrid, K., Cullen, P., & Haake, D. (2004). Cell aggregation: a mechanism of pathogenic *Leptospira* to survive in fresh water. Int Microbiol, 7(1), 35-40.

TSCL. (1987). Subcomittee on the Taxonomy of *Leptospira*. Int J Syst Bacteriol, 37, 472-473.

Tsuchimoto, M., Niikura, M., Ono, E., Kida, H., & Yanagawa, R. J. Z. f. B., (1984). Mikrobiologie und Hygiene. Series A: Medical Microbiology, Infectious Diseases, Virology, Parasitology. Leptospiral attachment to cultured cells. 258(2-3), 268-274.

Tubiana, S., Mikulski, M., Becam, J., Lacassin, F., Lefèvre, P., Gourinat, A.-C., . . . D'Ortenzio, E. J. P. n. t. d. (2013). Risk factors and predictors of severe leptospirosis in New Caledonia. 7(1), e1991.

Tungtrakanpoung R, Pitaksajjakul P, Na-Ngarm N, Chaicumpa W, Ekpo P, Saengjaruk P, et al. (2006). Mimotope of *Leptospira* from phage-displayed random peptide library is reactive with both monoclonal antibodies and patients' sera. Vet Microbiol. 115(1):54–63. https://doi.org/10.1016/j.vetmic.2006.02.011.

Turner, M. D., Nedjai, B., Hurst, T., & Pennington, D. J. (2014). Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 1843(11), 2563-2582. doi:https://doi.org/10.1016/j.bbamcr.2014.05.014

Tutar, Y. J. C., & genomics, f. (2012). Pseudogenes. 2012.

Urquhart, B. L., Cordwell, S. J., & Humphery-Smith, I. (1998). Comparison of Predicted and Observed Properties of Proteins Encoded in the Genome of *Mycobacterium Tuberculosis* H37Rv. Biochemical and Biophysical Research Communications, 253(1), 70-79. doi:https://doi.org/10.1006/bbrc.1998.9709

Victoria, B., Ahmed, A., Zuerner, R. L., Ahmed, N., Bulach, D. M., Quinteiro, J., & Hartskeerl, R. A. (2008). Conservation of the S10-spc-alpha locus within otherwise highly plastic genomes provides phylogenetic insight into the genus *Leptospira*. PLoS One, 3(7), e2752. doi:10.1371/journal.pone.0002752

Vijayachari, P., Sugunan, A. P., & Sehgal, S. C. (2001). Evaluation of microscopic agglutination test as a diagnostic tool during acute stage of leptospirosis in high & low endemic areas. *Indian J Med Res*, 114, 99-106.

Villumsen, S., Pedersen, R., Borre, M. B., Ahrens, P., Jensen, J. S., & Krogfelt, K. A. (2012). Novel TaqMan(R) PCR for detection of *Leptospira species* in urine and blood: pit-falls of in silico validation. J Microbiol Methods, 91(1), 184-190. doi:10.1016/j.mimet.2012.06.009

Vinetz, J. M., Glass, G. E., Flexner, C. E., Mueller, P., & Kaslow, D. C. J. A. o. i. m. (1996). Sporadic urban leptospirosis. 125(10), 794-798.

Viswanathan, V. K., Kurtz, S., Pedersen, L. L., Abu Kwaik, Y., Krcmarik, K., Mody, S., & Cianciotto, N. P. (2002). The Cytochrome c Maturation Locus of *Legionella pneumophila* Promotes Iron Assimilation and Intracellular Infection and Contains a Strain-Specific Insertion Sequence Element. 70(4), 1842-1852. doi:10.1128/IAI.70.4.1842-1852.2002 %J Infection and Immunity

VON HUNOLSTEIN, C., PARISI, L., TISSI, L., RECCHIA, S., ALFARONE, G., NICOLINI, L., . . . OREFICI, G. J. J. o. m. m. (1999). Virulence properties of type VII *Streptococcus agalactiae* (group B streptococci) and immunochemical analysis of capsular type polysaccharide. 48(11), 983-990.

Wagenaar, J. F., de Vries, P. J., & Hartskeerl, R. A. (2004). Leptospirosis with pulmonary hemorrhage, caused by a new strain of serovar Lai: Langkawi. J Travel Med, 11(6), 379-381.

Weil, A. (1886). Ueber einer eigenhuemliche, mit Milztumor, Icterus un Nephritis einhergehende, acute Infektionskrakheit. Deutsch Arch Klin Med, 39(209).

Werts, C., Tapping, R. I., Mathison, J. C., Chuang, T. H., Kravchenko, V., Saint Girons, I., . . . Ulevitch, R. J. (2001). Leptospiral lipopolysaccharide activates cells through a TLR2-dependent mechanism. Nat Immunol, 2(4), 346-352. doi:10.1038/86354

Wildschutte, H., Wolfe, D. M., Tamewitz, A., & Lawrence, J. G. J. P. o. t. N. A. o. S. (2004). Protozoan predation, diversifying selection, and the evolution of antigenic diversity in *Salmonella*. 101(29), 10644-10649.

Wolff, J. W., & Bohlander, H. J. (1966). A new serotype of *Leptospira*, belonging to the autumnalis serogroup. Trop Geogr Med, 18(3), 247-249.

Wolff. J, W., & Broom, J. C. (1954). The Genus *Leptospira Noguchi*, 1917. Problems of Classification and a Suggested System based on Antigenic Analysis. Documenta de Medicina Geographica et Tropica, 6(1), 78-95.

Wu, H., Soler-García, Á. A., Jerse, A. E. J. I., & immunity. (2009). A strain-specific catalase mutation and mutation of the metal-binding transporter gene mntC attenuate *Neisseria gonorrhoeae* in vivo but not by increasing susceptibility to oxidative killing by phagocytes. 77(3), 1091-1102.

Xu, Y., Zhu, Y., Wang, Y., Chang, Y. F., Zhang, Y., Jiang, X., . . . Wang, J. (2016). Whole genome sequencing revealed host adaptation-focused genomic plasticity of pathogenic *Leptospira*. Sci Rep, 6, 20020. doi:10.1038/srep20020

Yasuda, P. H., Steigerwalt, A. G., Sulzer, K. R., Kaufmann, A. F., Rogers, F., & Brenner, D. J. (1987). Deoxyribonucleic Acid Relatedness between Serogroups and Serovars in the Family Leptospiraceae with Proposals for Seven New *Leptospira* Species. 37(4), 407-415. doi:doi:10.1099/00207713-37-4-407

Yuri, K., Takamoto, Y., Okada, M., Hiramune, T., Kikuchi, N., & Yanagawa, R. (1993). Chemotaxis of leptospires to hemoglobin in relation to virulence. 61(5), 2270-2272.

Zaki MJ, Meira W (2014). Data Mining and Analysis: Fundamental Concepts and Algorithms: Cambridge University Press

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res, 18(5), 821-829. doi:10.1101/gr.074492.107

Zhang, L., Xie, J., Wang, X. e., Liu, X., Tang, X., Cao, R., . . . Liang, S. (2005). Proteomic analysis of mouse liver plasma membrane: Use of differential extraction to enrich hydrophobic membrane proteins. 5(17), 4510-4524. doi:doi:10.1002/pmic.200401318

Zhao, X., Norris, S. J., & Liu, J. (2014). Molecular Architecture of the Bacterial Flagellar Motor in Cells. Biochemistry, 53(27), 4323-4333. doi:10.1021/bi500059y

Zhong, Y., Chang, X., Cao, X.-J., Zhang, Y., Zheng, H., Zhu, Y., . . . Guo, X.-K. (2011). Comparative proteogenomic analysis of the *Leptospira interrogans* virulence-attenuated strain IPAV against the pathogenic strain 56601. Cell Research, 21, 1210. doi:10.1038/cr.2011.46. https://www.nature.com/articles/cr201146#supplementary-information

Zuerner, R. L. (1994). Nucleotide Sequence Analysis of IS1533 from *Leptospira borgpetersenii:* Identification and Expression of Two IS-Encoded Proteins. Plasmid, 31(1), 1-11. doi:https://doi.org/10.1006/plas.1994.1001

Zuerner, R. L., Alt, D., & Bolin, C. A. (1995). IS1533-based PCR assay for identification of *Leptospira interrogans* sensu lato serovars. 33(12), 3284-3289.

Zuerner, R. L., & Bolin, C. A. (1990). Nucleic acid probe characterizes *Leptospira interrogans* serovars by restriction fragment length polymorphisms. Vet Microbiol, 24(3), 355-366. doi:https://doi.org/10.1016/0378-1135(90)90183-V

Zuerner, R. L., Herrmann, J. L., & Saint Girons, I. (1993). Comparison of genetic maps for two *Leptospira interrogans* serovars provides evidence for two chromosomes and intraspecies heterogeneity. 175(17), 5445-5451. doi:10.1128/jb.175.17.5445-5451.1993 %J Journal of Bacteriology

Zuerner, R. L., & Huang, W. M. (2002). Analysis of a *Leptospira interrogans* locus containing DNA replication genes and a new IS, IS1502. FEMS Microbiol Lett, 215(2), 175-182. doi:10.1111/j.1574-6968.2002.tb11388.x

Zwietering, M., Jongenburger, I., Rombouts, F., Van't Riet, K. J. A., & microbiology, e. (1990). Modeling of the bacterial growth curve. 56(6), 1875-1881.

10.0 ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisors, **Assoc. Prof. Dr. Michael Hust** and **PD Dr. Frank Pessler** for their invaluable input, guidance and continuous support over the four years. Thank you for all the opportunities you have offered to me. My deepest gratitude also goes to **Dr. Boyke Bunk**, for being an excellent mentor in bioinformatics works, as well as for his support and contribution in the concept and progress of the research.

I would also like to thank, my collaborators; **Dr. Zbigniew Arent, Assoc. Prof. Dr. Andreas Beineke, Dr. Birgit Strommenger, Dr. Robert Geffers, Dr. Marga Goris, Dr. Cathrin Spröer, Dr. Michael Jarek**, **Dr. Sabin Bhuju, Dr. Natalia Novoselova, Dr Van Kihn Nguyen, Dr. Jonas Zantow** and **Gustavo Moreira** for their contributions in material, knowledge and networks. My appreciation also goes to **Assoc. Prof. Dr. Susanne Haeussler** and **Dr. Robert Geffers** for their valuble input and critical feeback in the thesis committee meetings.

This study would not have been possible without the crucial contribution of **Dr. Normaznah Yahya, Dr. Salawati Mustakim, Dr. Zahratul Laily Abdul Hadi, Dr. Hani Mat Hussin, Dr. Tajul Ariffin bin Awang Mohd,** who were so kind to provide precious material.

A big thank you to **Zulkifli Ubil, Hans van der Linden** and **Karin Littman-Janßen** for their technical assistance.

Finally, I would like to thank my family especially my three lovely daughters i.e. **Nur Alisya Sofea, Nur Arwin Insyirah** and **Nur Audrin Mardhiyah** for being patient and understanding throughout my PhD years. I dedicate this thesis to them!